Research Note 86-04

AD-A165 844

DTIC FILE COPY

COST AND INFORMATION EFFECTIVENESS ANALYSIS: AN IMPROVED METHODOLOGY
VOLUME 1

John K. Hawley, Bryan E. Brett, and William A. Chapman
Applied Science Associates, Inc.

for

ARI FIELD UNIT AT FORT BENNING, GEORGIA
Joel D. Schendel, Acting Chief

TRAINING RESEARCH LABORATORY
Seward Smith, Acting Director

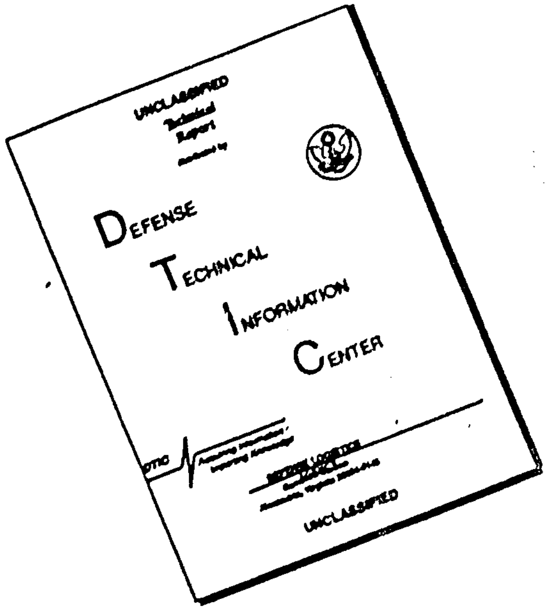
U. S. Army
Research Institute for the Behavioral and Social Sciences

January 1986

Approved for public release; distribution unlimited.

86 092

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

Applied Science Associates, Inc.

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
ARI Research Note 86-04

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
COST AND INFORMATION EFFECTIVENESS ANALYSIS: AN IMPROVED METHODOLOGY (VOLUME I)

5. TYPE OF REPORT & PERIOD COVERED
Final Report
Feb 1981 - June 1982

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
John K. Hawley, Bryan E. Brett and William A. Chapman

8. CONTRACT OR GRANT NUMBER(s)
MDA903-80-C-0212

9. PERFORMING ORGANIZATION NAME AND ADDRESS
APPLIED SCIENCE ASSOCIATES, INC.
Box 158
Valencia, PA 16059

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
2Q263743A794

11. CONTROLLING OFFICE NAME AND ADDRESS
US Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Ave., Alexandria, VA 22333

12. REPORT DATE
January 1986

13. NUMBER OF PAGES
208

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)
ARI Field Unit
P.O. Box 2086
Ft. Benning, GA 31905

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Seward Smith, Contracting Officer's Representative

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Training Device
Performance Assessment
Information Worth
Information Utility
Multiattribute Utility Measurement
Psychological Scaling
Combat Simulation Model
Cost-Benefit Analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents the results of an effort to develop an improved methodology for the conduct of Cost and Information Effectiveness Analysis (CIEA). CIEA is a methodology for the evaluation of training device performance assessment capabilities (D-PACs). It is directed at the problem of determining when the worth of performance status information available from a D-PAC offsets the costs required to develop, operate, and maintain the capability. Such a determination may be needed in order to specify requirements for a D-PAC, or as a basis for deciding between two or more D-PAC design

Cont.

20.

options intended to satisfy pre-specified requirements.

Following the introductory section, a review of objective procedures for the assessment of information worth is presented. An objective method for information worth evaluation based upon the use of Combat Simulation Models (CSMs) is then explored in detail. The results of this evaluation indicated that a CSM-based CIEA procedure, while technically feasible, is not practical.

~~Section 3 of~~ the report presents results from a series of formative tryouts of alternative multiattribute utility measurement (MAUM) procedures for the conduct of CIEA. Based upon these empirical results, recommendations for an improved MAUM-based CIEA methodology are made.

The report continues with a detailed presentation of an improved methodology for the conduct of CIEA. In this improved methodology, a series of MAUM procedures are integrated into a standard cost-effectiveness framework. To illustrate the methodological description, an exemplary analysis on a set of hypothetical D-PAC alternatives is included.

Finally, in section 5, issues relevant to the application of the improved CIEA methodology are discussed. Suggestions for future methodological development are also presented.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Cont'd)

1. INTRODUCTION

Most scenarios for a full-scale confrontation between the United States and its major potential adversaries indicate that the majority of Army units will have to be prepared to fight immediately without the luxury of a lengthy mobilization and "train-up" period. Studies of the comparative military strengths of the United States and its allies versus the Warsaw pact countries also indicate that friendly forces are likely to be heavily outnumbered. This potential situation has been termed a "come as you are war", with the added requirement of "winning the first battle outnumbered" (DuPuy & Gorman, 1978).

To have any hope of success in an engagement such as the one alluded to in the previous paragraph, the U.S. Army must maintain a high level of individual and unit combat readiness at all times. Maintaining a consistently high level of combat readiness will necessitate frequent and accurate evaluations of individual, crew, and unit proficiency, along with a means of quickly diagnosing and remediating indicated performance deficiencies. Increased training by itself is not sufficient. To be cost-effective, training, particularly in a unit setting, should be directed at specific problem areas; thus the need for an adequate performance assessment capability (PAC).

In the distant past, the frequent evaluation of individual, crew, or unit proficiency presented no special difficulties. Ordnance, POL (petroleum, oil, and lubricants), spare parts, and other support items were relatively inexpensive and readily available. As a result, live-fire training/evaluation exercises were held with sufficient frequency to provide commanders with a reasonable indication of their units' combat potential. Recently, however, the complex nature of many new weapons systems, the cost and limited availability of ordnance for these systems, and the cost of other support items have resulted in a situation in which live-fire exercises on a scale necessary to assess and maintain combat readiness are no longer feasible. Commanders are thus faced with the dilemma of knowing that if war comes they must be ready to fight immediately but not having the training/evaluation resources necessary to provide an expectation of success in such an engagement.

A partial solution to the problem of conducting more frequent proficiency evaluations in an era of increasingly restrictive resource constraints is the use of training devices (e.g., simulators, mockups, etc.) instead of actual equipment in the conduct of such evaluations (Finley, Gainer, & Muckler, 1974; Hopkins, 1975). In addition to their training applications, training devices can provide a vehicle for individual and collective performance assessment (Fitzpatrick & Morrison, 1971; Glaser & Klaus, 1972; Crawford & Brock, 1977). Historically, the most extensive uses of training devices in performance assessment have been in the aviation community (Caro, 1973). The commercial airlines and the Federal Aviation Administration currently use flight simulators extensively in aircrew performance certification. Follow-up studies have indicated that pilot performance in flight simulators is predictive of performance in actual aircraft (American Airlines, 1969; Weitzman, Fineberg, Gade, & Compton, 1979).

In a military setting, the uses of training devices in performance assessment have generally mirrored civilian applications and primarily involved aviation. There has been, however, an increasing use of training devices to assess individual and collective performance in other areas, such as maintenance (Hanson, Harris, & Ross, 1977) and anti-submarine warfare (Bell & Pickering, 1979; Callan, Kelley, & Nicotra, 1978). In the Army, one long-standing, non-aviation program of individual and collective performance assessment based upon the use of a training device is found in the Air Defense branch. Here, the AN/TPQ-29 engagement simulator is used in the conduct of a variety of performance evaluation exercises for HAWK missile personnel. The AN/TPQ-29 (and prior to that the AN/MPQ-T1 simulator on the Nike-Hercules system) is an engagement simulator capable of generating a variety of simulated air defense combat situations [e.g., multiple targets, electronic countermeasures (ECM) of various kinds, etc.]. The simulator was designed primarily for use as a training device, but it is also used to evaluate individual and crew performance. When using the AN/TPQ-29 in performance assessment, an evaluation team loads a "raid tape" containing the parameters of a simulated air defense mission into the HAWK system's computer. The HAWK crew is evaluated on its ability to defeat the simulated threat; performance checklists are used to evaluate individual crew members. Hardcopy printouts of some individual and crew performance measures (e.g., targets destroyed, numbers of penetrators, track engagement times, operator reaction times, etc.) are also available from the computer.

The evaluation of aircrew members in a flight simulator or HAWK personnel using the AN/TPQ-29 illustrates the concept of a training device performance assessment capability, or D-PAC. The term D-PAC simply means that a proficiency assessment capability is included with the training devices for a materiel system. Once built into the training device system, the D-PAC is used to assess the job proficiency of the individuals or crews that operate the materiel system.

A recent review of Army training device proficiency assessment potential indicated that the D-PAC principle can be applied to the training devices for virtually any materiel system (Shelnutt, Smillie, & Bercos, 1978). At the present time, actual use as in the aviation community or in HAWK air defense units is not widespread, but the potential remains. In one sense, D-PAC implementation is implicit in the development of any training device. Realistically speaking, however, D-PAC implementation may require an extension or modification of a training device to provide information that is: (1) immediate, and (2) useful to commanders or trainers. The cost-effectiveness of a D-PAC may thus vary as a function of the extent of training device modifications and extensions versus the payoff resulting from the receipt of additional proficiency status information. In this same sense, cost-effectiveness is also a function of the incremental payoff of D-PAC information compared to the cost of obtaining the same information in other ways (e.g., through the conduct of live-fire exercises).

Given that the D-PAC concept has potential for application in virtually any training device system, a critical issue concerns the circumstances under which such capabilities should be developed. That is, determining the conditions under which the proficiency status information available from a projected D-PAC is worth the cost of developing and operating the training device extensions and evaluation system modifications required to obtain and make use of that information. Such a cost versus information benefit analysis has been termed Cost and Information Effectiveness Analysis, or CIEA.

Research Background

The U.S. Army Research Institute (ARI) Field Unit at Ft. Benning, Georgia, with contractor support from Applied Science Associates, Inc. (ASA), has initiated a research program concerned with evaluating the D-PAC concept and developing the means for its implementation in existing and emerging training device systems. One of the primary objectives of this research program has concerned the development of a methodology for the conduct of CIEA. During the first year of the research program, a preliminary CIEA methodology based upon the use of Multiattribute Utility Measurement (MAUM) and various other psychological scaling procedures was developed, tested, and evaluated (see Hawley & Dawdy, 1981a, 1981b).

Figure 1-1 presents a block diagram of the major steps in the application of the preliminary CIEA methodology. The process begins with the definition of D-PAC objectives and constraints. The major product of the objectives/constraints phases is the identification of information worth dimensions (WDs), or major usage categories for D-PAC produced proficiency status information. Examples of typical WDs include:

1. Unit readiness evaluation
2. Unit training management
3. Unit management
4. Fighting system evaluation/development

Following these exploratory actions, the next step in the analysis is concerned with the specification of D-PAC operational requirements. For the job position(s) under consideration, performances, conditions and standards are identified. Also as part of step 3, performances are operationally defined in terms of observables (i.e., cues, responses, reaction times, processes,

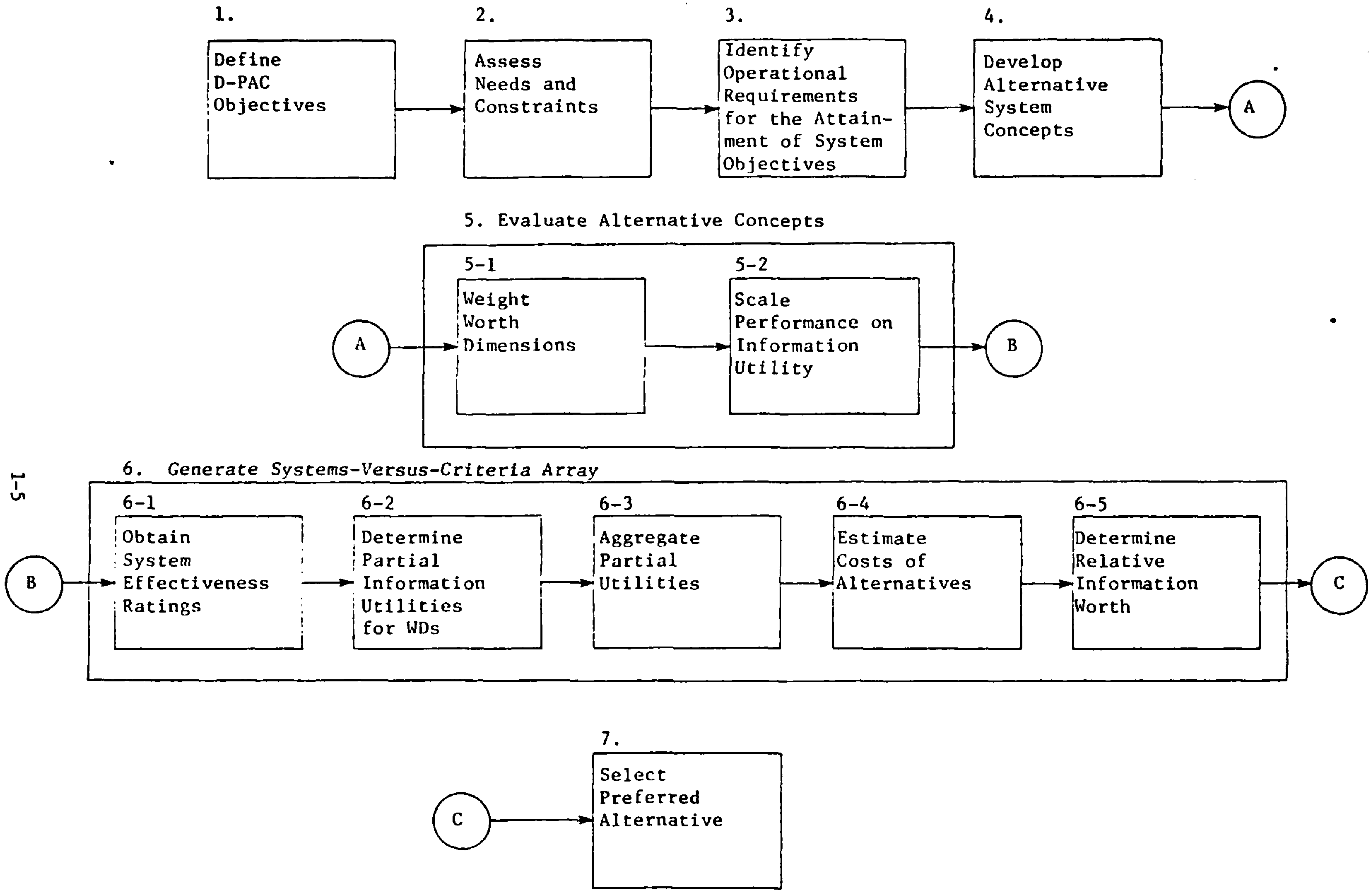


Figure 1-1. Preliminary CIEA Methodology

products, etc.) within the job environment; that is, operational performance measures (OPMs) are established.

Step 4 is concerned with characterizing alternative D-PAC concepts that meet the operational requirements set down in step 3. This portion of the procedure currently consists of integrating one or more training devices or performance evaluation vehicles into a set of D-PAC alternatives. A complete delineation of D-PAC alternatives includes:

1. Hardware requirements (device[†] specifications, numbers of devices, facilities, acquisition schedules, replacement rates, etc.).
2. Performance assessment methods.
3. A usage scenario (frequency of evaluation, evaluator requirements, expected length of evaluation period, etc.).

The D-PAC alternatives are specified at a level of detail sufficient to permit life-cycle cost estimates (LCCEs) to be developed.

Once D-PAC alternatives have been defined, the second activity in phase 4 involves the construction of a Performance by Alternatives matrix. Entries in this array are either a "1" or a "0" indicating, respectively, that D-PAC alternatives do or do not permit assessment of specific performances. Next, each cell containing a "1" (i.e., performance assessment is possible) is elaborated upon through an explicit consideration of the performance assessment method used and the devices' coverage of target/condition variables [referred to hereafter as performance context variables (PCVs)]. Each assessment method is rated according to the judged precision of the data it provides. Factors that are considered in assigning precision ratings include both reliability (i.e., stability upon replication) and the content validity of the OPM.

[†]The term "device", in this context, denotes either a training device per se or an equivalent performance assessment vehicle, such as Record Fire on the M16A1 rifle.

Step 5 begins the process of evaluating the alternative D-PAC concepts. As a first activity, the performances identified in step 3 are mapped to individual WDs. This action is taken because it is recognized that not all performances are relevant to all WDs; that is, the value of information on specific performances for specific applications is judged a priori to be zero.

The second action in step 5 is to assign the WDs weights that reflect their importance relative to the D-PAC objectives established in step 1. Following this action, the last sub-step in step 5 addresses the worth of performance status information vis a vis each of the WDs. As currently structured, the CIEA information worth (IW) evaluation method is based upon the application of a riskless MAUM technique. Subject matter experts (SMEs) are guided through a scaling process intended to elicit numerical values reflecting the relative worth of performance status information for the applications subsumed under each of the WDs. Since a MAUM procedure is used to establish information worth, the results are necessarily subjective in nature.

After the generation of information utility scores for performances, the next set of activities in the CIEA concern the development of the systems-versus-criteria array. In CIEA, as in orthodox cost-effectiveness analysis (see, for example, Kazanowski, 1968), the systems-versus-criteria array is a matrix that explicitly presents each D-PAC alternative and its associated evaluation criteria. First, information quality (IQ) ratings are obtained for each D-PAC alternative on each performance. In this context, IQ is defined in terms of measurement precision (MP) and coverage of relevant contextual variables. Quality ratings are assigned holistically on a 0-to-100 scale using an anchored, direct subjective estimate (DSE) scaling procedure (see Torgerson, 1958).

The second aspect in the procedure is to determine the utility of receiving performance status information at the frequencies associated with the various D-PAC alternatives. In this regard, a DSE scaling procedure is used to obtain frequency utility (FU) ratings. As a third substep, the IQ and FU ratings are combined to form a single measure of effectiveness for each alternative on each performance (block 6-1).

The second substep in step 6 involves the computation of partial information utility (PIU) scores for each alternative on each WD. Expression (1-1) gives the rule for aggregating effectiveness scores across performances to yield PIU ratings:

$$PIU_{ij} = \sum_k U_{jk} E_{ijk}. \qquad (1\text{-}1)$$

In (1-1), PIU_{ij} is the PIU score of the i^{th} D-PAC alternative on the j^{th} WD (i.e., usage category);

U_{jk} is the utility score of the k^{th} performance nested under the j^{th} WD;

and E_{ijk} is the effectiveness score of the i^{th} D-PAC alternative for the k^{th} performance nested under the j^{th} WD.

PIU scores are next combined across WDs (block 6-3) to obtain an overall information utility (IU) score for each D-PAC alternative:

$$IU_i = \sum_j W_j \, PIU_{ij} , \qquad (1\text{-}2)$$

where IU_i represents the aggregate IU score for the i^{th} D-PAC alternative;

W_j is the importance weight of the j^{th} WD:

and PIU_{ij} is the partial IU score of the i^{th} alternative on the j^{th} WD.

The IU scores for D-PAC alternatives generated in this fashion represent the benefit measure for the CIEA. These measures reflect: (1) the extent and judged quality (i.e., reliability and content validity) of the data provided by each D-PAC alternative; (2) the judged utility of the D-PAC evaluation frequencies, (3) the judged relative worth of status information on the performances under consideration; and (4) the relative worth of each of the potential applications for the proficiency status information.

Step 6 continues with the development of LCDEs for D-PAC alternatives. In costing D-PAC alternatives, only those costs uniquely associated with performance evaluation are included. Costs associated with developing and procuring the constituent devices for training purposes are not included in the D-PAC LCCEs.

The last activity in step 6 concerns the actual construction of the Systems-Versus-Criteria array. At a minimum, this matrix displays IU scores and LCCEs by D-PAC alternatives. If the assumption that IU scores follow at least an equal-interval scale is judged to be tenable (see Torgerson, 1958), then the systems-versus-criteria array can be expanded to include relative information utility (RIU), relative information cost (RIC), and relative information worth (RIW). RIU is obtained by dividing the IU measure for each D-PAC alternative by that of the "base line" alternative:

$$RIU_i = IU_i/IU_b. \tag{1-3}$$

The baseline alternative is either the presently used or most conventional D-PAC alternative. RIC is determined in a similar fashion:

$$RIV_i = LCC_i/LCC_b\ , \tag{1-4}$$

where LCC_i is the life-cycle cost of the i^{th} alternative and LCC_b is the life-cycle cost of the baseline option.

In order to identify the most cost-effective D-PAC alternative, RIU and RIC can be combined to form RIW:

$$RIW_i = RIU_i/RIC_i\ . \tag{1-5}$$

An RIW score greater than one indicates that alternative i is more cost-effective than the baseline option. What is done, in effect, is to normalize the figures of merit for cost and utility, with the baseline alternative assigned a unit value. Again, it should be noted that the consideration of RIU, and thus RIW, is warranted only if the equal-interval scaling assumption for IU is met.

The final aspect of the CIEA is to select a preferred D-PAC alternative (step 7). If the equal-interval scaling assumption for IU is judged tenable (thus justifying the use of RIU and RIW), the decision-rule is simple: maximize RIW. If RIU and RIW are not judged appropriate for use in identifying a preferred D-PAC alternative, then the analyst must subjectively integrate IU with cost to select a preferred option.

Overview of Research Extensions

The objective of the IU assessment portion of the CIEA methodology is to provide a measure of the value for decision-making of performance status information obtainable from a D-PAC. Stevens (1951) defines measurement as a process of assigning numbers to objects or events according to rules. The preliminary, MAUM-based CIEA methodology outlined in the previous paragraphs defines one set of rules, or one "yardstick", for measuring information worth. However, the suitability of the MAUM-based yardstick has not been established. Prior to widespread application, it is desirable to demonstrate that: (1) the MAUM-based CIEA procedure is broadly generalizable; and (2) the extent to which resulting IU scores are (a) reliable, (b) properly scaled, and (c) predictively valid indices of actual IW.

The primary requirement for a CIEA methodology is that it be broadly generalizable; that is, that the procedure be usable with training devices of differing complexity and at various stages in their developmental cycle. The preliminary CIEA methodology described herein has been exercised using a fielded training device system having only low to moderate complexity (see Hawley & Dawdy, 1981b). Hence, the first methodological issue to be addressed in a research extension concerns the methodology's applicability in more complex situations.

Once it has been established that the overall methodological concept is operationally generalizable, a second methodological concern is the reliability of the process. In general, reliability refers to the consistency from one set of measurements to another on repetition of a measurement procedure (Stanley, 1971). In the case of CIEA, reliability denotes the stability or reproducibility of results upon repeated application of the procedure by independent users. A desirable state of affairs is that CIEA results be independent of whomever carries out the analysis.

A third methodological issue concerns the scaling properties of the IU scores resulting from an application of the procedure. In determining IU using the DSE scaling procedures, it is assumed that users are able to assign ratings following an equal-interval subjective scale. If this assumption is justified, then procedures similar to those used in the preliminary CIEA methodology provide scale values that have equal-interval properties. It should be noted, however, that the scaling methods themselves provide no explicit means of testing this assumption.

The assumption that decision-makers are capable of providing equal-interval scale values is critical to the system evaluation procedures currently used in CIEA. As noted earlier, the use of MAUM-derived IU scores in the computation of RIU and RIW is based upon the assumption that the level of measurement for IU is at least equal-interval (i.e., equal-interval or ratio). The effects on RIU and RIW (and thus the eventual selection of a preferred D-PAC alternative) of violations of the equal-interval assumption are not known. A reasonable conclusion, however, is that if IU is at most ordinal, then RIU and RIW are at most ordinal. The use of cost-effectiveness ratios (e.g., RIW) is based on the assumption that both the numerator and denominator terms are at least equal-interval. Using this means for integrating cost and effectiveness measures is inappropriate if the scaling properties of either numerator or denominator are suspect. In fact, the potential undesirable effects of violations of the equal-interval assumption have resulted in a general aversion to the use of cost-effectiveness ratios within the military systems analysis community.

A fourth methodological issue, somewhat related to the issue of the scaling properties of IU scores, concerns overall methodological complexity. The preliminary CIEA methodology employs a mixture of decomposition and holistic utility rating procedures to determine IU. Decisions concerning the use of decomposition versus holistic judgments at various places in the procedure were made on the basis of previous, and sometimes conflicting, research and applications, and on the basis of perceived limits on the complexity of the resulting analytical method. Since these decisions were often judgmental and arbitrary (on the part of the designers of the process), the suitability of the scaling procedures used throughout the analysis should be examined. The intent of this examination is to refine the methodology by using decomposition methods where they are most appropriate and holistic methods where they are most appropriate.

A desirable aspect of CIEA is that the methodology be as simple as possible while still producing acceptable results.

A fifth methodological issue, relevant to the viability of the entire MAUM-based CIEA process, concerns the validity of IU results, that is, establishing that IU does, in fact, mirror actual information worth, or IW. Given the current state of CIEA methodological development, the validity of IU will likely have to be tested within what is termed a convergent validation framework (Campbell & Fiske, 1959). In the present situation, convergent validation will involve determining the extent to which IU is related to other independently derived measures of IW. Hence, the first requirement in validating the use of IU in CIEA will involve developing one or more independent, non-MAUM-based methods for assessing IW.

Given the utility of a workable CIEA methodology in the development of cost-effective D-PACs, ARI desired to continue the development and application of the CIEA methodology initiated during the first year of the research effort, with an emphasis on the methodological issues noted above. Specifically, the emphases of the second year's effort include: (1) the refinement and validation of the MAUM-based CIEA methodology, and (2) the exploration of alternative means (i.e., not based on MAUM) techniques for developing measures of the worth of performance status information. The desired end-product of the study is an improved CIEA methodology retaining the best features of the old, but incorporating new procedures where old methods are found to be deficient.

In this regard, sections 2 and 3 of the report present the results of a series of analytical/empirical studies directed at the five methodological issues cited above. To begin the discussion, section 2 is concerned with the identification, review, and analysis of alternative IW evaluation procedures. Section 3 presents results from a series of formative evaluations directed at refining the MAUM-based CIEA methodology. The results from both of these sections are integrated to form the basis for an improved, MAUM-based CIEA procedure.

The topic of section 4 is an improved CIEA methodology that incorporates the results of the formative studies described in sections 2 and 3. Section 4 presents a detailed description of the improved CIEA methodology, along with an exemplary analysis intended to illustrate the logic and application of the method. Finally, in section 5, the discussion of CIEA methodological developments is terminated with a review of outstanding application issues. This material is followed by a series of general recommendations for potential users of the methodology. Suggestions for additional CIEA methodological developments and refinements are also presented and discussed.

2. METHODOLOGICAL DEVELOPMENT I: ALTERNATIVE INFORMATION WORTH EVALUATION PROCEDURES

As noted in section 1, the objective of CIEA is to assess the relative worth of the differential information obtainable from alternative D-PACs that have different costs and other associated resource requirements. The rationale for CIEA is that performance status information has value only when it results in gain to a receiving party. Hence, performance status information should be collected only when the cost of the collection effort is offset by a gain realized through the information's receipt. Implied in the objective and rationale for CIEA is a requirement to measure a construct denoted herein as information worth, or IW. The reader should recall that under the preliminary CIEA methodology, IW is measured indirectly by deriving a quantity denoted information utility using a MAUM procedure. One of the primary objectives of the current effort is to explore IW evaluation methods that are more objectively based than MAUM. The topic of this section of the report is the results of various attempts at identifying a viable alternative to MAUM for use in assessing IW in CIEA.

Assessing Information Worth

Prior to considering methodologies for assessing IW, it is instructive to define what is meant by the term "information". Bedford and Onsi (1966) define information as "data evaluated for a specific use." Reviewing information is characterized as a process of ignorance reduction. The function of information is to reduce the amount and range of uncertainty under which decisions are made.

Two attributes are typically associated with information: amount and worth. The amount of information in a communication is determined by the reduction in uncertainty resulting from its receipt. Information amount can be assessed formally (although somewhat tediously) through the application of measures such as Shannon's H (Shannon, 1948; Shannon & Weaver, 1963).

IW has been defined formally as a function of changes resulting from the use of information in the pursuit of specific purposes (Bedford & Onsi, 1966). It is obvious that amount and worth are not necessarily correlated. For example, a communication may contain a large amount of information without being valuable in the sense of saying something useful to a recipient. IW is measured by a receiver in terms of the information's uses in decision making. Strictly speaking, this point of view implies that IW must be determined by evaluating the potential actions of decision makers before and after the receipt of a given quantity of information. Worth is determined from the expected gain resulting from taking one course of action versus another after receiving information. For example, by training on task X instead of on task Y, or by training P hours this week instead of Q hours (Bedford & Onsi, 1966; Thiel, 1967; Lev, 1969). This view of the IW evaluation process has been termed the Expected Value of Perfect Information (EVPI) approach (for example, see Hillier & Lieberman, 1980).

Determining IW through an EVPI approach as described above would seem to be a relatively straightforward procedure. In the case of CIEA, however, some potentially serious problems are apparent. First of all, the logistics of the worth evaluation process would likely be unmanageable. The alternatives of a range of decision makers would have to be considered, the differential costs (or gains) of numerous before and after decision scenarios would have to be estimated, the costs associated with some action differentials might be difficult or impossible to quantify, and so on. Based upon a review and analysis of such potential problems, it was determined that an EVPI approach to IW assessment would not prove to be a practical substitute for MAUM in the conduct of CIEA.

A second approach to the measurement of IW involves defining worth in terms of the effectiveness production function (EPF) relating specific performances, denoted p_i, to individual or collective combat effectiveness (CE):

$$CE = f(p_1, p_2, p_3, \ldots, p_q). \qquad (2\text{-}1)$$

If it were possible to determine the form of the EPF, then a logical case can be made for defining IW in terms of the contribution of each of the p_i to CE; that is,

$$IW_i = g(b_i), \tag{2-2}$$

where IW_i denotes the worth of status information on performance $\underline{i}$,

b_i represents the relative contribution of p_i to CE,

and $g(\cdot)$ denotes a function relating b_i to IV_i.

The reader should note that the situation presented here represents a simplification of the IW assessment process. It should serve, however, to illustrate the general thrust of the argument being developed.

Following the notion presented above, the problem of measuring IW is reduced to specifying appropriate forms for expressions (2-1) and (2-2). Since determining an acceptable form for (2-2) is likely to be a lesser problem [i.e., $g(\cdot)$ can be defined in terms of normalized parameters from (3-1)], the ensuing discussion focuses upon potential ways of determining the form of (2-1).

As a point of departure, it can be stated with some certainty that the likely form for the EPF (2-1) will be a high-order polynomial. Furthermore, parameters in the estimated EPF will have to be determined by observing the effects on CE of changes in the p_i across a range of representative situations and then applying least-squares estimation techniques. In other words, to estimate the EPF, it will be necessary to obtain a number of controlled replications of individual or unit performance in combat or near-combat-like situations. Since obtaining controlled replications in an actual combat situation is obviously not feasible, the only practical means for achieving this end is a high-fidelity combat simulation.

High-fidelity combat simulations fall into two general categories: War Game Simulations (WGSs) and Combat Simulation Models (CSMs). WGSs range in complexity from board-games (i.e., the so called sand tables) to free-form gaming situations involving live exercises teams, referees, and a framework of scenarios. Most of the board-game WGSs are used primarily for training. Also, these simulations typically do not involve actual weapon use. Individual and/or crew performance usually is determined in a monte carlo fashion (Wood, 1981). Hence, board-game WGSs likely would not constitute a viable means for assessing IW in CIEA. One possible exception to this conclusion might be for selected aspects of command, control, and communication (C^3) performance.

Free-form gaming of the type that actually incorporates individual or crew performances is conducted periodically by the Army and other branches of the armed forces. For example, the combat simulations conducted by the Army at the National Training Center or by the Air Force in Operation Red Flag are representative of free-form war gaming. If sufficient controlled replications could be obtained, such simulations are capable of providing the data necessary to establish the form of the EPF. However, an overriding problem with free-form gaming is cost. A cursory review of the costs involved in the conduct of free-form exercises like Operation Red Flag indicates that such an approach is not practically feasible for consideration in CIEA. A possible exception to this conclusion involves the case of operator performance on certain missile systems (e.g., HAWK and PATRIOT). In these systems, the availability of environmental, full-task tactical operations simulators provides a capability for the conduct of high-fidelity, free-form exercises without the high costs associated with typical live exercises.

The second general class of high-fidelity combat simulations are the CSMs. In this context, the term CSM denotes a logical war-game model implemented on a digital computer. The use of a CSM permits a relatively rapid study of complex systems under varying conditions. If the CSM is a valid representation of the system under study then the results of the modeling effort can provide valuable insights into system capabilities, or can even be used to predict future system performance. In addition, since CSMs are employed using a digital computer, the cost of conducting the replicated exercises necessary to establish the form of the EPF is usually considerably less than the cost of comparable free-form exercises. CSMs thus appear, on the surface at least, to provide a feasible, objectively-based alternative to the use of MAUM in CIEA.

Combat Simulation Models

As noted in the previous paragraph, a CSM is a mathematical and logical combat model implemented on a digital computer. In order to clarify what actually constitutes a CSM, it is instructive to note first what is meant by the term model. Shannon (1975) defines a model as a "representation of an object, system, or idea in some form other than that of the entity itself." Models typically are constructed to facilitate studying a system in order to understand the relationships between its various components or to predict its performance under alternative operating policies. Actual experimentation with the real system may, however, be infeasible or cost ineffective. For this reason, system models are exercised as a surrogate for experimentation with the real system. Models are often implemented on a digital computer in order to facilitate obtaining the replications necessary to obtain stable estimates of various system performance indices.

Extending the above discussion, a CSM is a mathematial-logical representation of a combat situation developed for the purpose of studying the interplay of selected variables in that environment. CSMs have been developed and applied extensively within the armed forces [see the Department of Defense Catalog of Logistics Models (1980) for a review of representative CSM applications]. The attractiveness of CSMs in a military setting is generally attributable to three characteristics of simulation models (Shannon, 1975):

1. Simulation models permit a compression and/or expansion of time. Lengthy processes in the real world can often be compressed considerably using a CSM. Conversely, events that occur quickly in the real world can be explicitly expanded and decomposed to permit their study.
2. Variables that impact upon system performance can be controlled systematically. System performance can thus be studied without the confounding effects of uncontrolled concomitant variables.
3. CSMs permit experimental exercises to be replicated exactly. Results from one exercise using a model (CSM or other) must

be interpreted as random variables in a statistical sense (the same is probably true of the results from real-world or near-real-world exercises). The use of a CSM permits a study of the distribution of outcomes associated with a specific set of parameters. Combat can thus be studied as a probabilistic as opposed to a deterministic phenomenon.

Although CSMs are technically attractive and have been used to study a wide range of military problems, care must be taken that they are not employed indiscriminantly. Shannon (1975), Fishman (1978), and Law and Kelton (1982) present a series of general cautions regarding the use of simulation models that are applicable in the case of CSMs. First of all, in the application of a model, there is no guarantee of success. That is, a simulation can appear to reflect accurately a real world situation when, in actuality, it presents a biased picture. It is usually very difficult, if not impossible, to validate a large-scale model against real-world outcomes. As a result, model validation is typically treated at the level of face validity and reasonableness of output (i.e., the infamous Turing test). Along this same line, Shubik and Brewer (1972) note that the majority of simulation models "live off a very slender intellectual investment in fundamental knowledge". In other words, the parameters and relationships expressed in many models are based upon conjecture or assumption rather than upon actual experimental observation. This latter situation, when coupled with the inherent difficulty of validating large-scale models, suggests that caution be used when making decisions based upon model results alone.

A second caveat, somewhat related to the first, involves what Shannon (1975) has referred to as "deification of the numbers". Since the output of most simulation models consists of impressive arrays of numbers, there is a tendency on the part of users to accept model results without question. A model, especially when implemented on a computer, can thus assume an authority all out of proportion to what actually might be warranted.

A third caution concerns the use of simulation models to draw inferences or to predict beyond their intended range of application, without proper qualification. All simulation models (including CSMs) are developed to address specific and generally limited objectives (Law & Kelton, 1982). A considerable risk of arriving at improper conclusions is encountered when models are used for other than their intended applications. This caution is particularly relevant in the case of using CSMs in CIEA. The current generation of CSMs generally were not developed to explore the relationship between performance and combat effectiveness. Hence, using results from existing models to establish the form of the EPF would, in most instances, involve a potentially unsupportable application of model results.

CSMs in CIEA

The previous paragraphs presented a discussion of the rationale for the application of CSMs in a military setting. This portion of the report returns again to the primary thrust of discussion: an examination of alternative means of assessing IW in CIEA. However, the specific focus of the discussion is now directed at the applicability of using CSMs to this end.

In order to establish a concrete focus for the discussion, a CSM having potential for use in the conduct of CIEA has been selected for a case study. The CSM chosen to illustrate the potential application of CSMs in CIEA is the Army Small Army Requirements Study (ASARS) Battle Model. Using ASARS as a vehicle, a blueprint for the conduct of a CSM-based CIEA is presented. As the description of the analysis is developed, major issues relevant to the applicability and feasibility of the approach are identified. However, before continuing with the discussion of how a CSM-based CIEA would be conducted, it is instructive to provide a brief overview of the CSM used as a focus for the exemplary analysis: the ASARS Battle Model.

An Overview of the ASARS Battle Model

ASARS is a dynamic monte carlo simulation of two-sided, small unit dismounted combat. The model was originally developed to investigate the effects of weapons (i.e., small arms) performance characteristics such as ballistic dispersion, aim error, area coverage, rate of fire, lethality, and the like on engagement outcomes. Recently, however, the model has been employed to study integrated combat operations involving small unit fire and maneuver, with a mixture of small arms, mortar, and artillery fire, as well as firing from aircraft.

The design of ASARS is such that it can create a realistic representation of intense, small unit, close combat involving two sides with units ranging in size from fire teams (7 members) up through companies (simulated as the action of two platoons). Sub-models included in ASARS permit the specific consideration of:

1. Terrain features, plus the effects of vegetation type and density. Battlefield terrain features are represented in the model through the use of a digitized terrain map.
2. Up and down link communications. Communications directed to and received from an assumed higher command element as well as up and down within the basic units being simulated.
3. Unit movement (fire and maneuver). The model is sufficiently complex, or "intelligent", in its operation so that each side reacts "logically" to the actions of the other side. Unit reactions or movements are determined through a combination of decision rules and/or monte carlo (i.e., probabilistic) response selections.
4. Fire control. The capabilities of the two sides are controlled by setting firing rate and accuracy parameters. For example, with small arms fire, accuracy is varied through the specification of the standard deviation of the round dispersion pattern.
5. Casualty assessment. Casualty assessment is made on the basis of incapacitation as well as kills. The suppression effects of near misses are also considered.

ASARS provides no intrinsic measures of effectiveness (MOE). The model will, however, output whichever of 13 primary MOE are desired by a user.

Users must determine which MOE are desired and how they will be used. The 13 primary MOE provided by the model are listed as follows:

1. Measures of Supply Shortages. The number of times (cumulative) that each weapon is restricted from optimal employment due to ammunition supply levels.
2. Number of Rounds to First Hit. The average number of rounds fired by weapon type, per engagement of a specific target (area or point) to first hit.
3. Number of Hits per Burst. The number of hits achieved by a given weapon in a given engagement of a specific target within a preplanned and executed burst of fire.
4. Number of Different Targets Hit. The number of target elements--primary as well as others in the proximity--hit per engagement by weapon type.
5. Opening Engagement Range. The maximum range at which each weapon is employed against a suitable target.
6. Range at Which First Hit Occurred. The maximum range at which the first target hit occurred by weapon type.
7. Blue Casualties. The cumulative number of Blue casualties sustained.
8. Red Casualties. The cumulative number of Red casualties sustained.
9. Blue Casualty Rate. The time rate of casualty production.
10. Red Casualty Rate. The time rate of casualty production.
11. Percent of Time that Blue Maintains Fire Superiority. The percent (time/total time) of total time that Blue maintains fire superiority over Red. Fire superiority is defined as the greater relative volume of fire delivered into a target area. Unequal weapons (e.g., rifles vs. mortars) are combined in accordance with a "dangerousness" scale input to the simulation.
12. Red Suppression of Blue in Observation, Movement and Fire. The percent of total time that blue elements are suppressed

as a result of Red fire for observation, movement, and fire individually.

13. Blue Suppression of Red in Observation, Movement and Fire. The percent of total time that Red elements are suppressed as a result of Blue fire for observation, movement, and fire individually.

If an aggregate, or composite, MOE is desired, the user must specify the form of the combination function and compute it as an additional step in the simulation. The ASARS computer program is flexible enough, however, to readily permit the incorporation of additional code to achieve this end (i.e., model software is written in FORTRAN).

For additional detail regarding the structure or application of the ASARS Battle Model, the reader is referred to the documentation found in ASARS Battle Model: Executive Summary (1973) or ASARS Battle Model: Narrative Description (1973). These documents are the first and second volumes of a nine part series.

Blueprint for Application

As noted above, a description of ASARS is included herein because the model provides a potential vehicle for the conduct of a CSM-based CIEA. Specifically, the use of ASARS in the evaluation of a set of small arms D-PAC alternatives is considered. Small arms, in this context, is taken to mean the M16A1 rifle. In the next paragraphs, the conduct of a hypothetical CIEA evaluation of a series of M16A1 D-PAC alternatives using ASARS is described. This hypothetical evaluation parallels closely a similar demonstration analysis conducted on a set of D-PAC alternatives using the MAUM-based IW evaluation procedure (see Hawley & Dawdy, 1981b, or section 4 of this report).

Recall that one of the steps in the conduct of CIEA involves establishing the worth of performance status information for selected uses. These data, when integrated with ratings on the capabilities of each of the component devices, provide IU scores that are used as the effectiveness component of a cost-benefit type analysis. Under the MAUM-based method of analysis described in section 1 (and modified in section 4), IU scores are derived using a subjective evaluation procedure. SMEs are asked to rate the utility of receiving data on each of the performances. A structured psychological scaling procedure is employed in the elicitation of worth scores. In obtaining utility ratings on performances, it is assumed that information supplied to a commmander varies with respect to a single dimension of worth, denoted "utility". It is further assumed that it is possible to use scaled subjective estimates to measure this internal scale of worth. It is this scaled subjective aspect of IW evaluation that a CSM-based analysis would replace.

The actual conduct of the IW evaluation portion of a CSM-based approach to CIEA would likely involve the steps listed below. It should be noted that the steps listed below are tailored to an application of ASARS in CIEA. The application pattern for other CSMs should not be much different, however.

1. Establish Simulation Parameters:
    a. Set terrain/vegetation features.
    b. Define unit communication patterns.
    c. Define Red and Blue unit composition.
    d. Set unit movement parameters.
    e. Determine weapons mix for Red and Blue players.
    f. Decide what additional weapons will be employed (e.g., mortars, artillery, air, etc.).
    g. Establish weapon operating characteristics.
    h. Establish fire control doctrine (fire and maneuver characteristics for Red and Blue players).
2. Review MOE to determine indices of interest; define aggregation rule (if desired).
3. Map D-PAC performances to simulation independent variables.
4. Develop experimental design.
5. Conduct simulation runs.
6. Evaluate simulation results.
7. Apply results to the assessment of IW.

The first step in the application of ASARS in CIEA would be to establish the parameters used in the simulation; that is, to set the operating scenario. This would involve determining values for the parameters listed in (1-a) through (1-h) above. Some of these input variables will be used as independent variables in the simulation experiments; others will not be of direct experimental interest, but rather will define the context for the simulated engagements. Care must be taken to insure that the values selected result in a representative combat situation.

In many situations, setting the simulation parameters (i.e., defining the simulation data base) will be a lengthy, costly process. As an example, a review of the documentation for several typical CSMs indicates that the initial parameterization effort can take anywhere from four to six months to complete. In addition, a large amount of not-readily-available information such as digitized terrain maps is often required.

Following the establishment of simulation parameters, the second step in the conduct of a CSM-CIEA would involve reviewing ASARS' MOE regarding their relevancy to the D-PAC evaluation. It might also be judged desirable to aggregate several of the MOE to produce a summary measure of combat effectiveness. The nature of the aggregation rule would, of course, have to be determined subjectively.

After having identified appropriate MOE, the third step in the process would involve mapping relevant performances to ASARS' independent variables. For example, in the case of M16A1 D-PAC evaluation, those performances having to do with marksmanship proficiency would be mapped to the only simulation input variable pertaining to proficiency: the standard deviation of the circular round dispersion pattern. In many instances, the performance mapping process will be relatively gross; that is, many related performances will be mapped to one simulation variable, with no direct link between the individual performances and that simulation variable. ASARS, for example, does not treat behavior at a sufficiently molecular level to permit direct links between individual behavior and overall proficiency to be established. Situations will also be encountered in which entire clusters of tasks map to no simulation variable, or whatever mapping is developed is somewhat judgmental.

After the user has mapped system-relevant performances to simulation variables, the next step in the CSM-CIEA process would involve developing an experimental design suitable for use in establishing the EPF. Reviewing the simulation variables designated as independent variables (i.e., variables subject to experimental manipulation), factor levels must be defined and numbers of replications at each experimental level must be determined. The number of replications selected must insure that MOE are estimated with adequate precision.

The next step in a CSM-based CIEA process would be to actually conduct the simulation runs using a digital computer. Although this step is the crux of the analysis, it actually might turn out to be the most straightforward. In many instances, the user will not have to actually conduct the runs, rather they will be carried out by specialists at a computer center.

After receiving the simulation output, the user must next evaluate the results. This step would consist of using standard statistical procedures to establish the form of the EPF (most probably the application of a linear model approach). Once an acceptable form for the EPF is determined, the next step in the analysis would be to apply the EPF in the assessment of IW. If performances are mapped directly to simulation independent variables, then normalized parameters from the EPF or partial coefficients of determination (i.e., partial R^2 values) could be used for this purpose. If groups of performances are mapped to more global simulation variables, then the parameters from the EPF or normalized partial R^2 values could be used as constraints in a mixed MAUM-CSM analysis. In these cases, worth values for specific performances within groups can not be derived directly from the simulation results. However, "sub-EPFs", relating individual performances within task clusters to the worth of the overall cluster, could be determined subjectively using a method similar to the current MAUM-based procedure.

The above discussion raises several issues pertinent to the suitability of CSMs in the conduct of CIEA. Probably the most important point relates to Shannon's (1975) caution regarding the use of simulation models outside their intended realm of application. Most current CSMs (including ASARS) were not developed to address the objectives of CIEA (i.e., to establish the EPF). In application, this limitation will result in: (1) a non-correspondence between individual/collective performances and CSM independent variables, and/or (2) non-relevancy of simulation MOE. Compensating for these shortcomings would often result in a secondary application of MAUM procedures and thus an infusion of subjectivity into the CSM-based CIEA process. Recall that a desire to provide an objectively-based analytical procedure provided the impetus for the consideration of CSMs in the first place.

A second significant issue is the cost of a CSM-based approach to CIEA. Most of the CSMs currently available require large-scale support in terms of computer hardware and specialized expertise (e.g., systems analysts, programmers, operations research personnel, etc.). In addition, the costs associated with obtaining the necessary replications would invariable be high. For example, consider the case of using ASARS in the conduct of an M16A1 D-PAC evaluation. If three independent variables, each with five factor levels, were to be studied, the resulting experimental design would contain 15 cells, or simulation situations. Now, as a conservative estimate, suppose that 50 replications per simulation condition were required to achieve acceptable precision in the estimation of model parameters. This would require that 15x50 = 750 individual simulation runs be conducted. No figures concerning the cost of a single ASARS run were provided with the documentation for the model. Assume for purpose of exposition, however, that ASARS is roughly equivalent to CARMONETTE (another CSM) in terms of individual run cost (not an unreasonable assumption). The Defense Logistics Studies Information Exchange (DLSIE) estimates the average individual run time for CARMONETTE at 10 minutes (a range of 5-15 minutes) with an associated cost of $25. (In the opinion of the authors, $25 per run is a very conservative cost estimate.) If these estimates were applied to the case of using ASARS in an evaluation of a set of M16A1 D-PACs, an estimate of the total time required to conduct the required simulation runs is 750 x 10 minutes = 7500 minutes = 125 hours = 5.2 days. Assuming that $25 per run is a reasonable cost estimate, the total cost for the computer time associated with the analysis is $18,750.

Discussion

It appears technically feasible to apply, at least partially, some members of the current generation of CSMs in the conduct of CIEA. However, the application of CSMs not specifically designed to address the objectives of CIEA does pose some problems. The most notable problem is that varying degrees of subjectivity would remain in the analysis. Future generations of CSMs might permit a treatment of more molecular levels of individual or crew behavior. If this were the case, CSMs would be more suitable for use in CIEA.

The application of CSMs in CIEA, whether now or in the future, is likely to be a costly venture, however. This conclusion will be true in terms of time, special expertise, and direct outlays of money. In addition, if one were to consider the cost of developing CSMs specifically tailored to the objectives of CIEA, then the conduct of CIEA using CSMs would be a cost ineffective undertaking. An obvious conclusion in this regard is that the application of CSMs in CIEA, now or in the future, is likely to be an economically risky venture.

Given the limitations of the current generation of CSMs and the projected cost of developing and applying specifically tailored models, it must be concluded that the use of CSMs in CIEA, while technically feasible, is not practical. In the case of the current generation of models, much of the analysis would likely remain subjective in nature anyway. Hence, it is not clear that the results of a CSM-based procedure using current models would be superior to results obtained using a cheaper MAUM-based procedure. It is difficult to project the outcomes of CSM-based CIEAs using an improved generation of models. However, given the almost certainly high cost of developing and/or modifying and then exercising such models, the use of even improved models must be viewed as a potentially cost ineffective undertaking.

The review of alternative IW evaluation procedures presented in this section indicates that none of the methods considered is feasible or practical. This result has important implications for the development of a valid and reliable CIEA methodology. It suggests that CIEA will remain, for the present at least, dependent upon the application of a subjective MAUM-based procedure in the treatment of IW. The methodological developments considered in the next section of the report (i.e., concerning the refinement of the MAUM-based procedure) thus assume more importance than would have been the case if a practical alternative IW evaluation procedure were to have been identified.

3. METHODOLOGICAL DEVELOPMENT II: REFINEMENT OF THE MAUM-BASED CIEA PROCEDURE

As noted in section 1, the objective of the IW assessment portion of the CIEA methodology is to provide a measure of the value for decision-making of a given amount of information obtainable from a D-PAC. The MAUM-based CIEA procedure defines one means for assessing IW. The usability of the MAUM-based yardstick has, however, not been established. To date, little information regarding the actual properties of MAUM-based IU[1] scores has been produced. To be appropriate for use in CIEA, it is necessary to demonstrate that the MAUM procedure is broadly generalizable and that resulting IU scores are:

1. Reliable,
2. Properly scaled (i.e., at least equal-interval),
3. Predictively valid indices of strict (i.e., true) IW.

A primary requirements for any IW evaluation procedure is that it be broadly generalizable. Within the context of a D-PAC evaluation, generalizability refers to the methodology being usable with training devices of varying complexity (i.e., ranging from a few performances/conditions with a sophisticated measurement capability to a large number of performances/conditions with a sophisticated measurement capability) at various stages in their developmental cycle (e.g., conceptual, breadboard, fielded). The preliminary CIEA methodology developed during the first contract year was demonstrated using only fielded training devices having only low to moderate complexity, as indexed by the number of performance objectives relevant to

[1]It is customary in psychological scaling work to differentiate between the physical and the psychological continua. The physical continuum represents the true, but often unknown, scale of measurement for an attribute (e.g., length, brightness, pitch, information worth, etc.); the psychological continuum is the subjective representation of the physical continuum obtained through the application of various psychological scaling procedures. In the MAUM-based CIEA procedure, IW denotes the physical continuum, or true information worth. IU, obtained through the application of MAUM, denotes the psychological continuum, or judged worth.

D-PAC implementation and the sophistication of the associated measurement capability (see Hawley & Dawdy, 1981b). Hence, the first methodological issue to be addressed in a work extension involves a demonstration that the MAUM-based CIEA methodology is generalizable to a range of training devices. This issue can be resolved only through an application of the methodology to a series of training devices spanning the Device Complexity - Developmental Stage axes. The issue of procedural generalizability is not dealt with directly in this report. Rather, it is addressed separately as part of a full-scale demonstration of the improved MAUM-based CIEA methodology. The results of this demonstration exercise are reported in Brett, Chapman, and Hawley (1982).

Once it has been established that the MAUM-based CIEA procedure is operationally generalizable, a second methodological concern is the reliability of the method. In general, reliability refers to the consistency from one set of measurements to another on repetition of a measurement process (Stanley, 1971). In the case of CIEA, reliability denotes the stability, or theoretical reproducibility of IU results. An obviously desirable state of affairs is that CIEA results be reasonably independent of whomever constitutes the decision-making group, given that equally qualified representatives of the same stakeholders provide the constituent ratings.

Proceeding from this view of reliability, the reliability of the MAUM procedure will likely have to be assessed in a manner analogous to that of parallel forms reliability in psychological test theory. In psychological test construction, parallel forms reliability is established by first developing two independent testing procedures (i.e., parallel forms) assumed to provide the same true score. Next, each form of the test is administered to equivalent groups of testees. The correlation of results obtained using the two testing procedures provides the basis for computing a reliability coefficient (see Lord & Novick, 1968).

In the case of CIEA, establishing reliability must be done in a conceptually similar fashion: Independent groups of decision-makers representing the same stakeholders will complete the MAUM procedure evaluating the same set of D-PAC alternatives. The degree of consistency across groups will provide an indication of the reliability of the method.

Strictly speaking, the reliability assessment procedure outlined above will not demonstrate the absolute reliability of the MAUM-based CIEA method, but instead only the results of its application in a particular situation (e.g., for a training device of a given complexity at a given developmental stage and for a given group of decision-makers) (Torgerson, 1958). It would thus be desirable to demonstrate experimentally the reliability of the procedure across a range of devices to which the methodology might be applied. This can be accomplished through a replication of evaluation process across a range of training devices. In any event, even a single demonstration of reliability would serve to enhance user confidence in the results of the MAUM-based procedure.

A third methodological issue relevant to the usability of the MAUM-based IW evaluation procedure concerns the scaling properties of the ratings data. In assessing worth or value using a DSE scaling method, it is assumed that decision-makers are capable of rating various aspects of the D-PAC alternatives on an equal-interval subjective scale. If this assumption is correct, then the scaling procedures used in CIEA provide scale values that have equal-interval properties.

The assumption that decision-makers are capable of providing equal-interval scale values is critical to the system evaluation procedures currently used in CIEA. The use of MAUM-derived IU scores in the D-PAC evaluation is based upon the assumption that the level of measurement for IU is at least equal-interval (i.e., equal-interval or ratio). The effects of violations of the equal-interval assumption are unknown. However, the use of cost-effectiveness ratios is based on an assumption that both the numerator and denominator terms are at least equal-interval. Using this tool for integrating system cost and effectiveness measures is inappropriate if either the numerator or denominator terms are improperly scaled.

In view of its criticality for system evaluation, the validity of the equal-interval assumption should be examined empirically. As in the case of generalizability and reliability, testing the equal-interval assumption requires a repeated application of the MAUM-based procedure in equivalent evaluation situations. Repeated applications are necessary because data obtained in a single application provide no basis for determining whether or not decision-makers are judging on the basis of an equal-interval scale (Torgerson, 1958). It is always possible to compute scale values on the basis of an equal-interval assumption. In addition, the consistency of judgments across groups is, in itself, not an adequate criterion by which to evaluate the equal-interval assumption. Completely inconsistent judgments are evidence that ratings do not follow an equal-interval scale. Consistent ratings do not imply, however, that decision-makers are judging on the basis of an equal-interval scale. A criterion based on consistency alone does not distinguish between equal-interval judgments and simple ordinal position judgments.

The minimum requirement for an equal-interval scale is that the ratios of differences in scale values assigned to any three or more stimuli are invariant with respect to the values of the remaining stimuli in a set (Torgerson, 1958). This result can be verified experimentally by plotting the scale values obtained from one evaluation against the scale values from a second independent replication of the same set of stimuli. If the equal-interval assumption is met, the resulting plot will be linear, within sampling error. Again, as in the case of reliability, a demonstration that decision-makers used an equal-interval scale in one situation does not necessarily generalize to other situations. Repeated demonstrations of the validity of the equal-interval assumption will, however, build user confidence in the validity of using MAUM-derived IU scores in CIEA.

A fourth methodological issue relevant to the MAUM-basic CIEA procedure concerns overall methodological complexity. CIEA employs a mixture of decomposition and holistic utility evaluation procedures to determine IU. As noted earlier, decisions concerning the use of decomposition versus holistic judgments at various points in the current procedure were made on the basis of previous research and applications and on the basis of perceived limits on the complexity of the resulting analytical method. Should the preliminary version of the MAUM-based CIEA methodology not provide suitably scaled IU scores, a potential means of raising the level of measurement for IU is to examine the suitability of the utility evaluation procedures used in the analysis. The intent of the examination would be to refine the methodology by using the most appropriate evaluation procedures. That is, by using decomposition methods when they are most appropriate and holistic methods when they are most appropriate. The objective is to develop an operational procedure that is simple, face valid, theoretically correct, and reasonably robust in its application.

The issue of the validity of the MAUM-based procedure is not addressed in this section. In order to explore the validity issue, it is necessary, at a minimum, to have available one or more alternative measures of IW. The negative results of the efforts described in section 2 mean that for the present no alternatives to the MAUM-based CIEA procedure are viable, thus alternative measures of IW are not available. As a result, validity studies, other than those concerned with face and content validity, cannot be carried out.

The Developmental Studies

Given the unresolved issues noted above, the second major objective of the current effort concerned a systematic exploration of the methodological problem areas. This portion of the project assumes even more importance in light of the results presented and discussed in section 2. If CIEA is to be a viable methodology, then, for the present at least, it will be dependent upon the use of SME input. It is thus imperative that the means used to elicit and treat the data required to exercise the analysis be as sound as possible.

To begin the process of refining the preliminary CIEA methodology, the project staff first reviewed the old procedure with the objective of identifying and addressing obvious problem areas. Using results obtained during the first year of the effort and information derived from a further review of the cost-effectiveness and MAUM/psychological scaling literature, the basic methodological framework for CIEA was altered to make the procedure more logically consistent. The result of this restructuring exercise is shown as Figure 4-1 in the next section of the report. The reader should note that the improved methodological framework illustrated in Figure 4-1 does not represent a substantial departure from the preliminary framework depicted in Figure 1-1.

The review of the analytic procedure indicated five points in the analysis where SMEs input judgmental data. These points are listed as follows.

SME Input Point	Method
1. Worth Dimensions	Identification and weighting using a successive comparisons scaling procedure.
2. Information Utility	Ratings are obtained using a successive comparisons scaling procedure.
3. Measurement Precision	Holistic procedure using a DSE scaling approach.
4. Coverage of Performance Context Variables	Not treated explicitly. Considered with MP in assigning IQ ratings.
5. Frequency Utility	DSE rating on a 0-to-100 scale.

Also listed above are the methods employed in the preliminary CIEA methodology to elicit judgments from SMEs.

Working from the results of the preliminary review, the project staff next identified the most likely means for improving the quality of the analysis and/or lessening its methodological complexity. These rating "points of inquiry", so to speak, were also selected to provide a vehicle for studying the reliability of the procedure and the validity of the assumption that typical Army SMEs can provide rating data that follow an equal-interval scale. The rating points of inquiry selected for additional study are presented in Table 3-1. Table 3-1 also re-lists the current method used to elicit SME judgments and identifies one or more alternatives for each method.

Table 3-1

Points of Inquiry in MAUM-CIEA Formative Studies

Rating Point	Current Method	Alternative Methods	
1. Performance Utility	Successive comparisons	Method: Ranking Simple Rating Paired-Comparison	Structure: Hierarchical Inference with rollback vs. "Flat" approach
2. Measurement Precision	Direct Subjective Estimate (DSE) Holistic Approach	DSE with Decomposition Approach	
3. Performance Context	Non-Explicit Treatment	Explicit Treatment	

Note that rating point 1, Performance Utility, involves two separate issues. The first issue concerns the rating method used, and the second concerns the structure within which to employ the method. Currently, a "flat" rating structure is employed. That is, SMEs rate the entire set of performances as a single, undifferentiated group. In large scale applications, this so called flat approach can be confusing and cumbersome. An alternative to the flat approach is to place the performances in a hierarchical framework. That is, to develop a structure in which performances map to sub-functions and sub-functions to functions (see Figure 4-2 for an example of such a structure). Under this approach, SMEs assign ratings only within individual levels of the hierarchy. Utility scores are obtained by multiplying through the hierarchy, or "rolling back" the hierarchy, so to speak. Any of the rating methods can be used within this hierarchical inference (HI) structure.

The In-House Studies

During a discussion with the Contracting Officer's Representative (COR), the project staff was alerted to the possibility of not being able to obtain sufficient numbers of Army SMEs to study all of the alternatives noted in Table 3-1. As a result, a decision was made to conduct a series of preliminary evaluations using selected members of the project staff as test subjects. After some additional discussion with the COR, it was decided that these so-called "in-house studies" would address the issue of which of the IU rating methods to use in an improved version of the CIEA methodology. As noted in Table 3-1, three alternatives to the Successive Comparisons (SC) method were identified; these are:

1. Ranking
2. Simple Rating
3. Paired-Comparison (PC)

Each of the alternative rating methods has the associated benefit of being easier to employ than the SC procedure.

To evaluate each of the rating methods, a trial exercise was designed and conducted. This trial rating exercise employed project staff members working in a group setting. To lessen so-called "remembering" effects, the separate rating procedures were exercised on different days. The performances used as rating stimuli in the test exercise were specific to the STINGER Air Defense missile system. STINGER performances were chosen as test stimuli because all of the in-house participants were familiar with that system, and thus could be considered quasi-SMEs. Because the SC procedure was currently in use, it was selected as the standard against which to judge the other methods. Recall again that the objective of the in-house exercise was to determine which, if any, of the less complex rating methods could be used in place of the cumbersome SC method. Instructions for the application of each of the rating methods used in the in-house study are presented in Appendix A.

The correlations between the scale values resulting from the in-house exercise are presented in Table 3-2.

Table 3-2

Correlations Among Alternative Utility Rating Methods

	Rank	Rate	PC	SC
Rank	--			
Rate	.20	--		
PC	.99	.20	--	
SC	.79	.23	.82	--

The results indicate that the Ranking and PC procedures yielded roughly the same utility information as the SC method. There are, however, several problems with the Ranking and PC methods that limit their utility in application. First of all, the Ranking method makes the assumption that the stimuli being rated are uniformly distributed over the scale. In the case of CIEA, such an assumption would be difficult to defend. The PC method, while generally resulting in equal-interval scaled results, does so only when the number of stimuli being rated exceeds a threshold number (a commonly stated threshold number is seven). For smaller numbers of stimuli, the PC procedure results in an ordinal scale. The constraints imposed on the scale values by repeated comparisons within a group of stimuli produce an equal-interval scale.

Because of the limitations noted above, the project staff inquired into the suitability of a simple Rank and Rate procedure. This procedure is essentially the first step in the SC method. When the Rank and Rate procedure was independently applied to the STINGER performances, the resulting scale values correlated $\underline{r} = 0.83$ with the results obtained using the complete SC method. Since the Rank and Rate scaling method (1) has high face validity, (2) theoretically produces equal-interval scale values (see Johnson & Huber, 1977), and (3) produced results roughly equivalent to the standard SC procedure, it was selected as the method to be used to obtain utility scale values in the improved CIEA methodology.

The Ft. Benning Formative Tryouts

Having decided upon a preferred method with which to elicit utility scale values, the next step in the methodological investigation involved the conduct of an additional series of formative study to address the remaining points of inquiry listed in Table 3-1. More specifically, this next step included a study of: (1) whether to use a structured (i.e., HI) versus an unstructured approach to the elicitation of utility scores; (2) whether to use a holistic versus a decomposition approach to obtaining MP ratings; and (3) whether to address the issue of PCVs explicitly or non-explicitly, as in the preliminary CIEA methodology.

In an effort to resolve these issues, eight SME groups (designated A through H) were obtained for participation in the additonal series of formative tryouts. The SME groups each consisted of three persons; their source and composition is given as follows:

A. Platoon Leaders, 197th Infantry Brigade

B. USAIS DTD, System Development Branch

C. Company Commanders, 197th Infantry Brigade

D. USAIC Infantry Training Brigade (Officers)

E. Battalion S-3 Staff, 197th Infantry Brigade

F. USAIS Systems Analysis Division, WGMD

G. USAIC Individual Training Group (NCOs)

H. USAIC Infantry Training Brigade (NCOs)

In accord with the discussion presented in the introductory portion of this section, the eight SME groups were used to obtain replications of each of the procedures. The quasi-experimental layout used to obtain replicated test data is presented in Figure 3-1. Instructions for each of the methods used in the formative tryouts are given in Appendix B.

Utility Rating Structure. To study the issue of which utility rating structure to use, the test groups each provided performance utility IU scores for two WDs: Unit Readiness Evaluation and Unit Training Management. One set of utility ratings was obtained early in the exercise, while the second set was collected as a last step. To partially control for learning effects, fatigue, and the like, the order in which the two sets of ratings were obtained was counterbalanced. As noted on Figure 3-1, groups A, C, E, G employed the HI procedure with rollback (RB). Groups B, D, F, and H provided utility scores using the non-structured, or flat, rating procedure. All eight groups used the Rank and Rate method in providing utility scores.

The correlation matrices for the utility scale values provided by the eight groups on Unit Readiness Evaluation and Unit Training Management are presented as Tables 3-3 and 3-4, respectively. In terms of ease of application, the HI procedure was reportedly easier for the SMEs to apply. However, from a review of Tables 3-3 and 3-4, the HI procedure did not provide scale results that were more consistent across groups than did the more cumbersome flat rating technique. Neither rating procedure, in fact, produced consistent scale values across groups. There are several rather high correlations

Method/Procedure

Measurement Precision:	Decomposition DSE				Holistic DSE			
Performance Context Variables	Explicit Treatment		Non-Explicit Treatment		Explicit Treatment		Non-Explicit Treatment	
Utility Rating Structure	Hierarchical	Flat	Hierarchical	Flat	Hierarchical	Flat	Hierarchical	Flat
Group	Platoon Leaders, 197th Bde	USAIS DTD	Company Commanders, 197th Bde	USAIC ITB (OFF)	S-3 Staff, 197th Bde	USAIS SAD, WG MD	USAIC ITG (NCOs)	USAIC ITB (NCOs)
	A	B	C	D	E	F	G	H

Figure 3-1. Formative Evaluation Group Use Plan

Table 3-3

Correlations Among Utility Ratings
for Unit Readiness Evaluation

	B	D	F	H	A	C	E	G
B	1.0							
D	.047	1.0						
F	.677	.0236						
H	.534	.014	.336	1.0				
A	.350	.016	.716	-.004	1.0			
C	.052	-.2233	.1515	-.1711	.040	1.0		
E	-.086	-.490	-.025	-.141	-.136	.628	1.0	
G	-.074	-.401	-.158	-.184	-.338	.581	.772	1.0

Table 3-4

Correlations Among Utility Ratings
for Unit Training Management

	B	D	F	H	A	C	E	G
B	1.0							
D	.407	1.0						
F	-.233	.277	1.0					
H	.653	.699	-.012	1.0				
A	-.274	.059	.460	-.282	1.0			
C	.635	.531	.230	.764	-.123	1.0		
E	.147	-.014	-.227	.137	-.513	.159	1.0	
G	.379	.177	.011	.395	-.300	.468	-.283	1.0

in Tables 3-3 and 3-4, but a secondary review of the correlation patterns indicated no apparent reasons why some groups provided consistent utility scores while others did not.

Recall that the objective of this portion of the formative tryout was to determine which, if either, of the two rating procedures is preferable in terms of: (1) ease of application, (2) consistency of results (i.e., reliability), and (3) the tenability of the equal-interval assumption. In this regard, the results indicate that the hierarchical procedure is preferred in terms of ease of application. However, neither procedure produced results that were consistent across groups. From this latter result, it must also be concluded that the absolute tenability of the equal-interval assumption for utility scores is questionable. Recall that consistency is not a complete test of the equal-interval sssumption, but a lack of consistency indicates that raters are not following a common equal-interval scale.

In summary, the results cited above are supportive of the following conclusions concerning the performance utility scores produced using the MAUM-based procedure:

1. Different groups cannot be counted upon to provide consistent utility results. The utility scores that result will be highly sensitive to group composition, mind set, and attitude. Hence, care must be taken in actual analyses to elicit utilities data from participants selected for their knowledge of the materiel system undergoing analysis and having a good "feel" for the information applications (i.e., WDs) they are addressing.
2. Since consistent utility scores were not obtained, the absolute tenability of the equal-interval assumption for these scores is in doubt. Utility scores elicited from carefully selected applications may meet the equal interval test, but the scores may not be reflective of any <u>absolute</u> underlying value scale. A different SME group might, in all likelihood, provide significantly different results.

In terms of which structuring procedure to use in an improved version of the MAUM-based CIEA methodology, the three criteria noted previously support the use of the hierarchical approach. The hierarchical procedure performed no worse than the flat, unstructured approach, but was considerably easier for SMEs to apply.

Measurement Precision. The next tryout exercises concerned the issue of whether a holistic versus a decomposition approach to assessing MP should be used in CIEA. Under the holistic method, SMEs provide MP ratings considering both reliability and validity, but not explicitly. The decomposition approach, on the other hand, requires SMEs to rate reliability and validity separately; these individual results are then directly combined to form an MP score.

Tables 3-5 and 3-6 show the MP correlations for the two exemplary devices [(Record Fire (RF) and Weaponeer (WP)] used in the formative tryout. Recall that groups A, B, C, and D used an explicit, decomposition approach, while groups E, F, G, and H used a holistic rating method. The results indicate that the MP scale scores obtained using both methods are quire similar. There is also a great deal of consistency across groups employing the same procedure (i.e., decomposition or holistic). The inter-group correlations for RF are somewhat higher, on the average, than the results for WP, but it can be hypothesized that this result is an artifact of the SMEs being more familiar with RF than with the WP.

As an interesting aside, a further analysis of the MP scale correlations indicates that, in the decomposition procedure, the participating SMEs were not able to separate the concepts of reliability and validity. Reliability and validity scale scores correlated highly with each other, as well as with the holistic MP results.

Considering now the issue of whether SMEs were able to provide MP ratings that follow an equal interval scale, Figure 3-2 presents a plot of the MP scale scores for six randomly selected performanc s provided by groups E and H (overall r = .98) on RF. The introductory paragraphs made the point that correlation, or consistency, alone is not sufficient to demonstrate that

Table 3-5

MP Correlation Results for Record Fire

		A	B	C	D	E	F	G	H
	A	--							
	B	.92	--						
	C	.87	.94	--					
Group	D	.88	.98	.93	--				
	E	.89	.98	.97	.98	--			
	F	.86	.84	.83	.83	.81	--		
	G	.86	.95	.95	.93	.96	.74	--	
	H	.82	.94	.94	.96	.98	.81	.91	--

Table 3-6

MP Correlation Results for Weaponeer

		A	B	C	D	E	F	G	H
	A	--							
	B	.84	--						
	C	.75	.95	--					
Group	D	.87	.91	.89	--				
	E	.66	.84	.96	.83	--			
	F	.87	.80	.79	.75	.76	--		
	G	.72	.66	.69	.61	.67	.79	--	
	H	.83	.87	.84	.86	.77	.74	.79	--

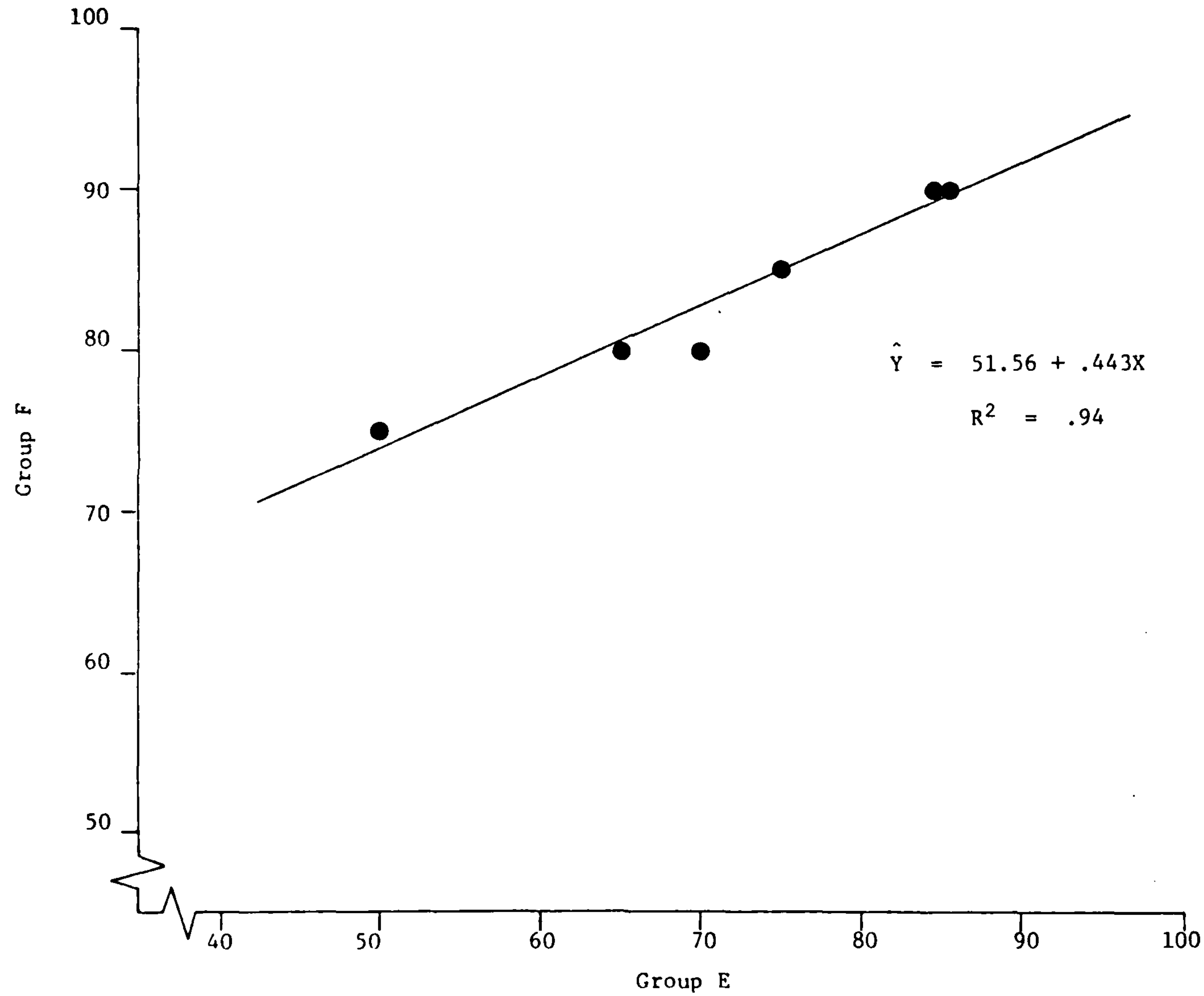


Figure 3-2. Plot of RF MP Scale Scores for Randomly Selected Non-Zero Entries

subjects are rating on an equal-interval scale. What is required is that a plot of the scale values of three or more stimuli obtained from two independent replications be linear. A linear regression analysis of the six scale point pairs resulted in $R^2 = 0.94$. A multiple correlation coefficient this high leaves very little margin for lack of fit to a simple linear model. In addition, the plot with the estimated regression line superimposed shows little dispersion of the points about the regression line. Admittedly, the results are not as striking in all of the cases observed, but the evidence presented above is encouraging in that it indicates that equal-interval scale values can be obtained using a DSE scaling procedure.

In view of the results presented above, a decision was made to opt for the holistic method in obtaining MP ratings. The results obtained using the holistic procedure were virtually identical to those obtained using the explicit decomposition method, but the holistic procedure is considerably easier for SMEs to apply.

Performance Context Variables. The last of the issues to be addressed in the formative tryouts concerned the treatment of PCVs. Recall that the preliminary CIEA methodology requires SMEs to consider MP together with what is termed "coverage of context variables" to provide a measure denoted as IQ. IQ is then integrated with FU yielding an Effectiveness score for D-PAC alternatives.

In the improved methodology, a decision was made to integrate MP with a PCV coverage rating, as before; the result is then integrated with FU to again form an Effectiveness score. The question at issue in the formative evaluations is whether PCVs should be considered explicitly in a decomposition framework, or non-explicitly as in the preliminary methodology.

To address the issue noted above, the formative evaluation groups were also divided on their treatment of PCVs. Groups C, D, G, and H used the older non-explicit, or holistic, approach. The correlations for the IQ ratings provided by these groups are presented as Tables 3-7 and 3-8.

Table 3-7

IQ Rating Intercorrelations for RF

	C	D	G	H
C	--			
D	.98	--		
G	.99	.97	--	
H	.98	.96	.99	--

Table 3-8

IQ Rating Intercorrelations for WP

	C	D	G	H
C	--			
D	.95	--		
G	.74	.81	--	
H	.85	.90	.86	--

The correlation patterns indicate that the test SMEs were able to provide consistent IQ ratings.

Consider, however, the correlations between IQ and the MP results for the same groups presented in Tables 3-9 and 3-10.

Table 3-9

IQ and MP Correlations for RF

	C	D	G	H
C	.97	.98	.98	.98
D	.82	.85	.80	.79
G	.97	.95	.93	.91
H	.93	.94	.98	.99

Table 3-10

IQ and MP Correlations for WP

	C	D	G	H
C	.64	.82	.82	.83
D	.80	.80	.82	.81
G	.66	.68	.71	.67
H	.86	.89	.90	.89

Since IQ is a component of MP, moderately high correlations between the two sets of scale values are to be expected. The magnitudes of the correlations obtained suggest, however, that the MP ratings are dominated by the IQ ratings. That is, that the SMEs actually may not have considered the contextual variables in any real sense when using the non-explicit rating procedure. If that were the case, then the MP ratings would not be sensitive to differences in device PCV coverage capabilities, as is desired. Such a result would render the holistic treatment of context variables unacceptable for use in an improved version of the CIEA methodology.

Such a possibility being the case, consider the exercises involving an explicit treatment of context variables. As part of the explicit procedure, SMEs are asked to rate relevant PCVs on their importance for inclusion in a D-PAC. These importance ratings provide the basis for the explicit decomposition evaluation procedure. In the formative tryout, groups A, B, E, and F were asked to provide importance ratings for a set of PCVs relevant to M16A1 marksmanship performance assessment (see Appendix A for a list of the PCVs used as stimuli). The correlations among the importance ratings obtained in the trial exercise are given in Table 3-11. These correlations represent a situation similar to that found with the utility ratings. That is, a great deal of variability in correlations is evident across groups. This result suggests that the SMEs were not able to assign PCV importance ratings in a consistent fashion. As with the utility ratings, different groups of SMEs will likely provide differing importance ratings for PCVs. The ratings will likely vary as a function of the SMEs' experience and current working perspective.

Table 3-11

PCV Importance Rating Correlations

	A	B	E	F
A	--			
B	.54	--		
E	.63	.47	--	
F	.04	-.09	-.23	--

In summary, the results pertaining to the choice of a method for treating PCVs are inconclusive. The non-explicit rating procedure provided consistent results, but there is evidence to suggest that the obtained consistency merely reflected an underlying consistency in MP ratings. On the other hand, the explicit, decomposition rating procedure provided inconsistent results across groups. However, after reviewing the results from the formative tryout and considering other relevant issues such as the face and content validity of the two PCV ratings procedures, a decision was made to employ the explicit decomposition approach to the treatment of context variables in the improved CIEA methodology.

Discussion

The results of the formative tryouts described herein are reflected in the procedures used in the improved CIEA methodology described in the next section of the report. Besides indicating the most appropriate methods to use at various points in the analysis, the formative tryouts are also revealing in another sense. The tryout findings indicate that CIEA results cannot be expected to be invariant across user groups. Different groups of users will, in all likelihood, provide differing evaluation results. Furthermore, the tenability of the equal-interval scaling assumption is in doubt, particularly for less-well-defined stimulus categories such as performance utility. Taken together, these results suggest: (1) that the cost-effectiveness type evaluation procedures (i.e., cost-effectiveness ratios) not be used in the improved CIEA methodology; and (2) that the CIEA procedure be viewed as a decision aid rather than as a mechanistic procedure for selecting a preferred D-PAC alternative. The formative tryout results indicate that the scaling procedures employed in the analysis are not sufficiently robust to support the latter level of application.

4. AN IMPROVED CIEA METHODOLOGY

As noted throughout the report, the objective of CIEA is to provide a framework for selecting a preferred D-PAC alternative in terms of the worth of the performance status information provided versus the cost of developing, implementing, and operating the capability. CIEA is intended to serve as a guide to decision-makers in establishing requirements for and in evaluating alternative D-PAC concepts for both fielded and emerging materiel systems.

As a methodology, CIEA is a member of a set of procedures generally known as cost-effectiveness analysis.[†] The term cost-effectiveness denotes a procedure in which alternative systems designed to meet specified goals are evaluated using measures of cost and separate measures of systems effectiveness (Barish & Kaplan, 1978). Under this approach, cost and effectiveness values for each alternative are determined. The systems are then evaluated on the basis of whether the incremental benefits of the more effective alternatives are worth their added costs. Cost-effectiveness type analyses are common in the evaluation of military materiel and training systems (for example, see TRADOC Pamphlet 11-8 or TRADOC Pamphlet 71-10).

Like its predecessor, the improved CIEA methodology described in this section of the report is developed within the framework of a general cost-effectiveness procedure outlined in Kazanowski (1968). The phases, steps, and major substeps of the methodology are listed as follows:

- 1.0.0 Concept Exploration
  - 1.1.0 Define D-PAC Objectives
  - 1.2.0 Assess Constraints

[†] Technically, CIEA as outlined in this section is a cost-benefit type analysis since a number of effectiveness measures are condensed into a single measure that serves as the basis for the evaluation of alternatives.

2.0.0 Concept Development

2.1.0 Define and Weight Information Worth Dimensions (WDs)

2.2.0 Define Performances, Conditions, and Standards

2.3.0 Map Performances to WDs

2.4.0 Define Operational Performance Measures

2.5.0 Specify Relevant Performance Context Variables

2.6.0 Obtain Priorities Data on Performances

2.7.0 Establish Utility of Performance Status Information for Selected Uses

3.0.0 Concept Definition

3.1.0 Define Hardware/Facility Requirements

2.3.0 Determine Performance Assessment Requirements/Methods

4.0.0 Concept Evaluation

4.1.0 Obtain Device Capabilities Matrix

4.2.0 Obtain Measurement Precision Ratings

4.3.0 Obtain Performance Context Matrix

4.3.1 Obtain Context Variable Importance Vector

4.3.2 Form Device Coverage Incidence Matrix

4.3.3 Form Absolute Coverage Matrix

4.3.4 Obtain Performance Relevancy Matrix

4.3.5 Compute Normalization Constants

4.3.6 Form Relevant Coverage Matrix

4.3.7 Form Performance Context Matrix

4.4.0 Form Device Measurement Effectiveness Matrix

4.5.0 Form Alternative Measurement Effectiveness Matrix

4.6.0 Obtain Frequency Utility Ratings for Performance Domains

4.7.0 Form Alternative Effectiveness Matrix

4.8.0 Form Partial Information Utilities Matrix

4.9.0 Compute Information Utilities

4.10.0 Estimate Life-Cycle Costs of D-PAC Alternatives

4.11.0 Summarize Results in Alternative-versus-criteria Array

4.12.0 Determine Most Cost and Information Effective Alternative

5.0.0 Design Specifications

5.1.0 Develop Detailed Design Specifications for Preferred Alternative

5.2.0 Develop Concept Validation Plan

Figure 4-1 graphically depicts the major elements in the improved methodology. Each of the phases, steps, and substeps of the analysis is now discussed in turn. To aid in the explication of the procedure, an exemplary analysis on a set of hypothetical D-PAC alternatives for the M16A1 rifle (the boxed-in sections) is presented with the narrative description. It is recommended that readers not familiar with the preliminary CIEA approach presented in Hawley and Dawdy (1981a) survey that material prior to proceeding into the current section.

Concept Exploration

Define D-PAC Objectives

Phase 1 of the CIEA process concerns establishing the need for a D-PAC, defining the objectives of the capability, and identifying general constraints that will serve as guide for the analysis. For fielded systems, the impetus for D-PAC development will come from a leadership concern that performance on a given materiel system is deficient. This concern may arise from a number of sources. For example, it may stem from low Skill Qualification Test (SQT) results, reports of poor Army Training and Evaluation Program (ARTEP) performance, or results from other individual or collective training/evaluation exercises. In other situations, the judgment that "things are not right" may be based on commanders' subjective opinions. In yet other cases, ammunition and other costs and constraints (e.g., availability) may limit the frequency with which performance status information is available, thus suggesting the need for an alternative to live-fire training/evaluation exercises. Whatever the source or basis, the impetus for the consideration of a D-PAC will be generated by the identification or perception of performance deficiencies that are judged to have significant impact on the Army's fighting ability.

For materiel systems under development, it is anticipated that CIEA will be conducted routinely as part of the Cost and Training Effectiveness Analyses (CTEAs) that accompany the training and training device development process.

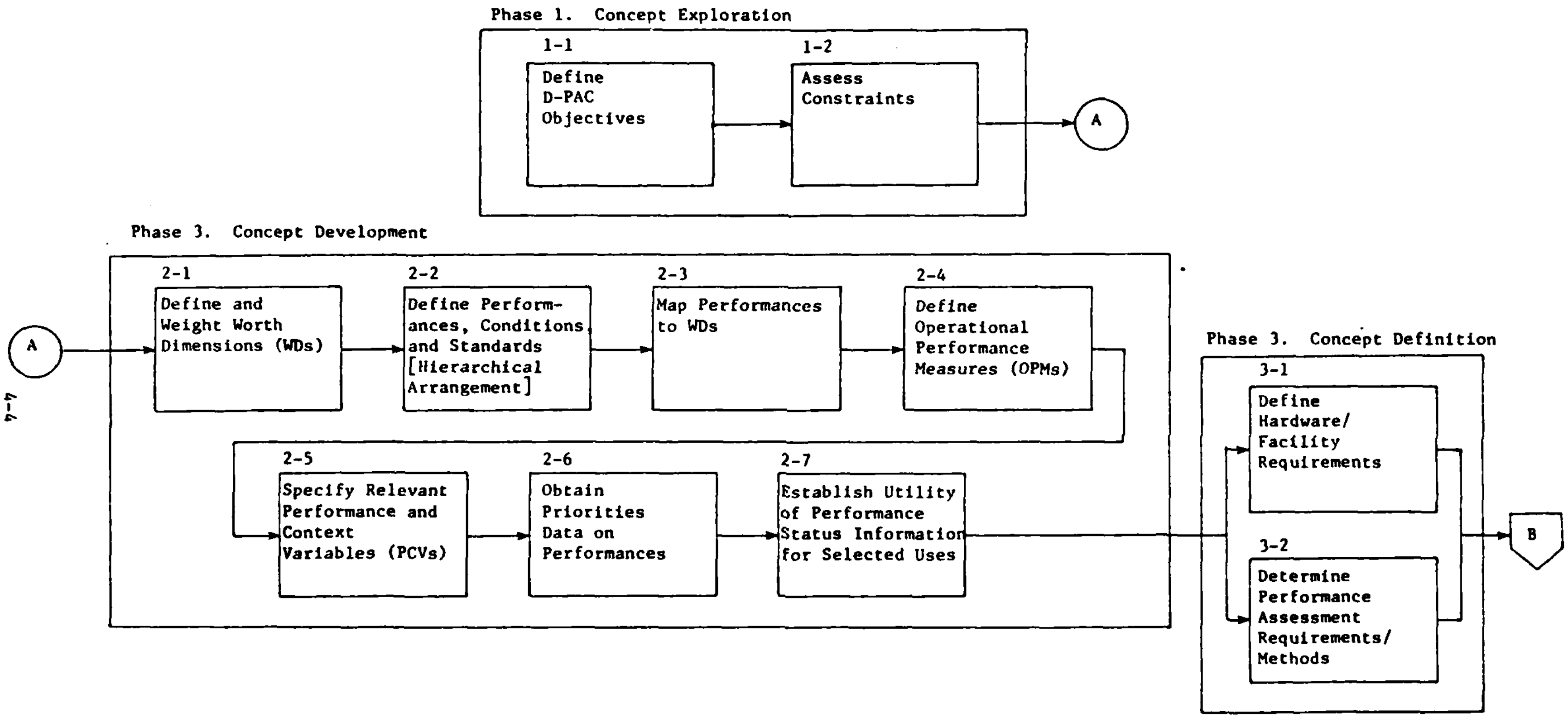


Figure 4-1. Improved Cost and Information Effectiveness Analysis

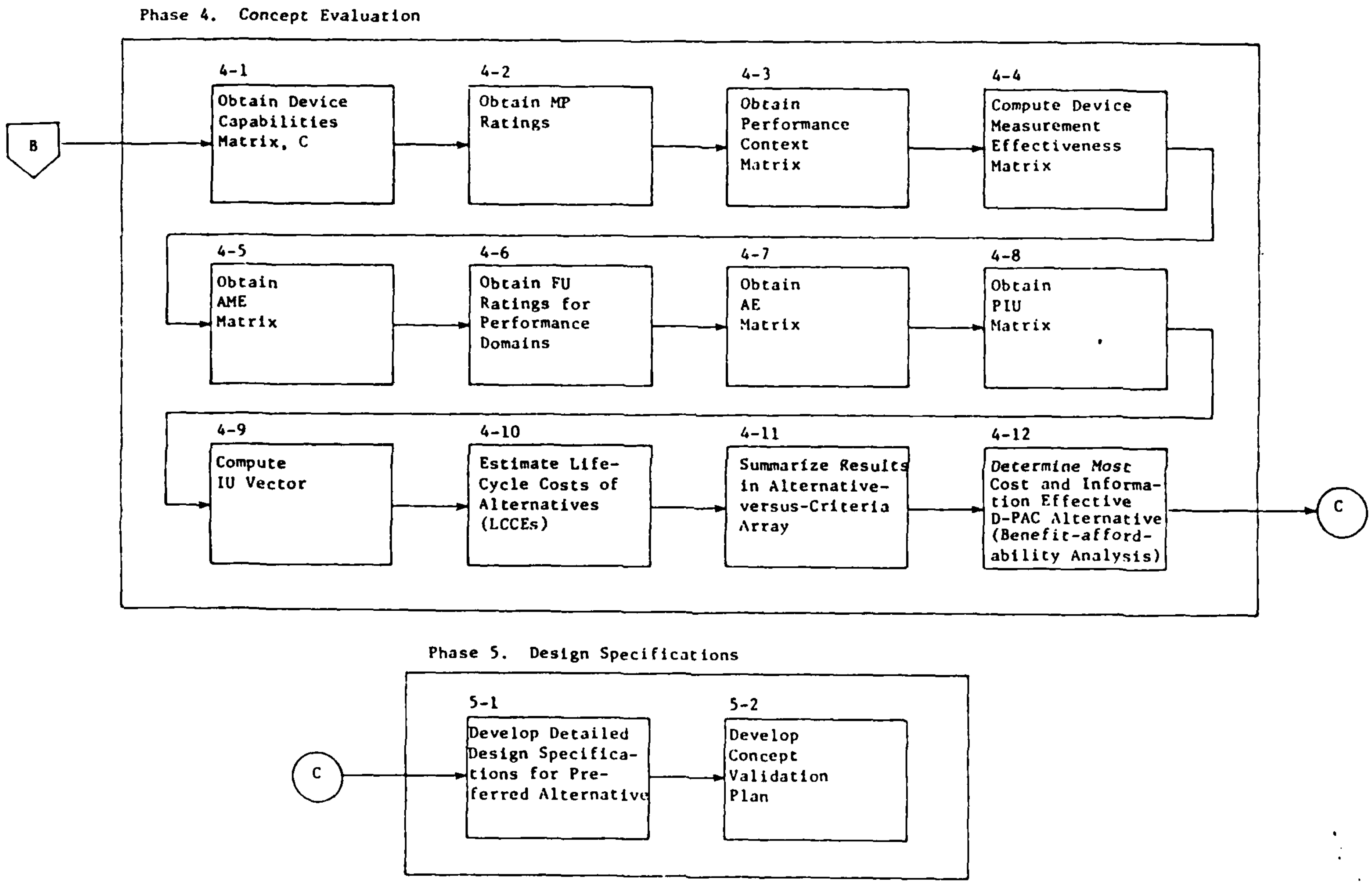


Figure 4-1 (Cont'd)

In these cases, the concern is in the area of potential performance problems that could be alleviated by more timely and effective performance assessment. Clues to potential high payoff D-PAC candidates in this regard could emerge from a review of performance problems on similar or antecedent materiel systems.

After the need for a D-PAC is established, it is necessary to formally set down the objectives of the capability. The objectives statement should define, in general terms, the basis for concern regarding performance deficiencies and identify the materiel system components and job position(s) involved. The objectives statement thus serves to define the initial range of performances to be considered in the developmental effort.

Assess Constraints

D-PAC development must be done in a real world situation in which constraints exist. Thus, after setting the general objectives for a D-PAC, the second step under Concept Exploration is to identify potential constraints on the development or deployment of the capability. Categories of constraints that may prove relevant include, but are not limited to, the following:

1. Economic
2. Technological
3. Personnel Requirement (quantity and quality)
4. Development Timeframe

It is doubtful that all applicable constraints can be identified early in the analysis. However, applicable categories of constraints should be identified. One constraint that should be addressed early-on is system cost. Cost will usually constrain the types of D-PACs that are developed and deployed. Hence, benchmark cost guidelines should be developed early in the analysis. Early determination of cost constraints will serve to eliminate excessively costly alternatives early in their developmental cycle.

Concept Development

Phase 2, Concept Development, is concerned with translating the general objectives statement produced in phase 1 into specific operational requirements for a D-PAC. This phase is carried out in six steps, described in the following paragraphs.

Define and Weight Information Worth Dimensions

Step 1 of phase 2 involves answering the question: "What purposes are the D-PAC-generated data to serve?" This question is answered by developing a list of information worth dimensions (WDs), or major categories of information use. The WDs constitute the primary value dimensions for the evaluation of D-PAC alternatives. Examples of some potential D-PAC WDs are listed as follows:

1. Readiness Evaluation. The determination of whether or not individuals/units are capable of performance at an acceptable level/standard on performances specific to the D-PAC implementation.
2. Training Management. The use of training status and performance diagnostic information in determining who, how often, when, and what to train for individual/unit performances related to the specific D-PAC.
3. Unit Management. The use of objective job performance information to provide guidance in various unit management activities such as the award of performance incentives, the assignment of personnel to critical unit positions, and so forth.
4. Fighting System Evaluation/Development. The use of evaluation data to provide feedback to branch schools and other concerned agencies on training program content, training materials, training devices, system equipment, support equipment, doctrine, tactics, and so forth.

In developing WDs it is necessary not to be too expansive. For reasons that will be apparent later, the number of WDs should not exceed seven. If more than seven WDs are developed, the list should be reviewed and the number of WDs reduced by redefining and combining dimensions. For most applications, the WDs listed above should suffice.

Following the specification of WDs, the dimensions are assigned weights reflecting their importance relative to the D-PAC objectives established in phase 1. Weights are assigned using the rank and rate scaling procedure described in Appendix C. The resulting weights range between zero and one, with the constraint that they sum to one.

After deriving importance weights for WDs, users should review the weights associated with specific WDs. Dimensions that have extremely low weights, relative to the others, should be eliminated from consideration. The elimination of relatively unimportant information usage categories at this point will significantly lessen the complexity of later stages of the analysis.

Define Performances, Conditions, and Standards

The second step in phase 2 concerns the specification of D-PAC operational requirements. For the job position(s) under consideration, performances (i.e., task statements), conditions and standards are defined. In most situations involving fielded materiel systems, existing task analysis documentation should provide the information necessary to develop performance statements, conditions, and standards (i.e., performance objectives). Situations may be encountered, however, (e.g., when working with an unfielded materiel system) in which performance objectives are missing or incomplete. In these cases, performance objectives will have to be developed by the analyst using: (1) a knowledge of antecedent systems; (2) preliminary materiel system documentation [e.g., the Logistics Support Analysis Record (LSAR)]; or (3) judgments rendered by SMEs.

Phase 2 continues with the development of a performance hierarchy (step 2.2.0). In this context, the term performance hierarchy denotes an arrangement that maps performances into sub-functions, and sub-functions into functions, or performance domains. An exemplary performance hierarchy for the M16A1 is presented as Figure 4-2. Users are encouraged to arrange performance in a hierarchical manner in order to facilitate applying the MAUM-based information worth evaluation procedure used later in the analysis.

Map Performances to Information Worth Dimensions

The third block of activities in phase 2 involves mapping performances to WDs. This action is taken because it is recognized that all performances may not be relevant to all WDs. In other words, it is judged a priori that information concerning particular performances for specific purposes is of no value. Removing non-relevant performances from the evaluation process at this point also has the effect of reducing the later complexity of the analysis.

Define Operational Performance Measures

After relevant D-PAC performances are identified and mapped to WDs, the next requirement in the analysis is to operationally define each performance in terms of observables (i.e., cues, responses, reaction times, processes, products, etc.) within the engagement environment. In CIEA terminology, these operationally defined performance statements are referred to as operational performance measures (OPMs). It may also be necessary in some situations to define OPMs for the sub-function level, the performance domain level, or even higher (i.e., for total, or aggregate, performance). Whatever the level at which performance assessment is required, it is necessary to specify exactly how individual/collective performance is to be characterized and quantified.

Specify Relevant Performance Context Variables

The fifth step in phase 2 concerns the specification of relevant PCVs. Context variables are environmental factors (e.g., discriminative stimuli,

Function/Performance Domain	Sub-Function	Performance
I. Prepare to Engage Targets	1. Condition Weapon to Operate	a. Calibrate weapon
		b. Load weapon
	2. Maintain Weapon Operation	c. Reload weapon
		d. Reduce stoppage
	3. Operate Weapon	e. Soldier-weapon interface
		f. Marksmanship factors
II. Engage Targets Individually	4. Assault Mode (using)	g. Aimed fire
		h. Pointed fire
	5. Defensive Position (using)	i. Aimed fire
		j. Pointed fire
	6. Patrol Operation (using)	k. Aimed fire
		l. Pointed fire
	7. Detect and Identify Targets	
III. Team Fire	8. Conduct Fire and Maneuver	m. Control fire team movement
		n. Move as a member of fire team
		o. Move under direct fire
	9. Practice Ammunition Rationing	p. Control rate of fire
		q. Control distribution of fire

Figure 4-2. Exemplary Performance Hierarchy for M16A1 Rifle

condition variables, target characteristics, etc.) that are judged to be significant moderators of job performance. For the convenience of potential users, a list of representative context variables is provided with this report as Appendix D. Note that at this stage of the analysis users are asked only to identify relevant context variables. The treatment of these variables in the analysis of D-PAC alternatives is addressed later during phase 4, Concept Evaluation.

Obtain Priorities Data on Performances

Following the specification of contextual variables, users are next asked to provide priorities data for the performances under consideration. Using the five criticality factors listed in Table 4-1, ratings are obtained for each performance. Then, after the five factors are arranged in the order of their importance (note: this is done independently for each application), the performances themselves are sorted into descending order of job criticality based upon SME responses to each of the factors. Under the current method of analysis, the job performance criticality rankings are not intended to drive the IW evaluation process. Rather, these data are obtained to provide a job context perspective for the IW ratings.

Establish Utility of Performance Status Information for Selected Applications

The final step in phase 2 concerns the worth of performance status information vis a vis each of the WDs. Following the results presented in section 3, the IW evaluation procedure used in the improved methodology is based upon the application of a hierarchical MAUM rating method. Using the instructions provided in Appendix E, SMEs are guided through the MAUM scaling process. The results of the scaling process are a set of numerical values reflecting the relative worth of status information on each performance for each WD.

Table 4-1

Performance Prioritization Factors

1. Consequence of Inadequate Performance - how serious is the effect of improper performance or non-performance on the unit or individual mission:
   - L = Has little or no effect on mission of individual or unit
   - M = Could degrade or delay mission performance
   - H = Could result in mission failure

2. Task Importance - is the task important to the survival of personnel and equipment?
   - L = Failure or non-performance would have little or no effect on survival of personnel or equipment
   - M = Failure or non-performance could endanger personnel or equipment
   - H = Task must be performed for survival of personnel or equipment

3. Time Delay Tolerance - what is the time allowed between receiving the task cue and starting the performance?
   - L = No need to start task at any specific time
   - M = Task start can be delayed for several minutes to a few hours
   - H = Must begin immediately or within a few minutes after cue

4. Frequency of Performance - how often is the task called for?
   - L = Infrequently - once a month or less
   - M = Moderate frequency - once every one to three weeks
   - H = Frequently - more often than once a week

5. Task Decay Rate - how frequently must the task be performed to assure that skills are not reduced below task standards?
   - L = Task skills require little or no practice to retain
   - M = Task requires infrequent practice - once every one to three months
   - H = Frequent practice required - more of ten than once a month

At this point in the discussion, the exemplary CIEA is introduced. The methodological illustration is structured around the development and evaluation of a D-PAC for the M16A1 rifle. For purposes of analysis, phase 1 (Concept Exploration) is assumed. It is further assumed that two WDs have been defined and weighted, as follows (step 2.1.0):

Worth Dimenstion	Importance Weight, W_j
Unit Readiness Evaluation (RE)	.33
Unit Training Management (TM)	.67

The job performances of interest are listed below (step 2.2.0):

Performance

I. Prepare to Engage Targets

1.	Conditon to operate:	a.	calibrate weapon
		b.	load weapon
2.	Maintain operation:	a.	reload
		b.	reduce stoppage
3.	Operate weapon:	a.	soldier-weapon interface
		b.	marksmanship factors

II. Engage Targets Individually in

4.	Assault mode using	aimed fire
5.	Defensive position using	aimed fire
6.	Patrol operation using	aimed fire

Note that the job performances listed above are a subset of those listed in Figure 4-2. Performance conditions and standards are not particularly relevant to the methodological illustration, thus they are not explicitly considered. Also, for the example exercise, all performances are judged relevant to both WDs (step 2.3.0). OPMs for the performances of interest (step 2.4.0) are given in the following table.

Performance			OPM		
I. Prepare to Engage					
1. Condition to operate:	calibrate weapon	go-no go	in accord with SOP		
	load weapon	"	"	"	
2. Maintain operation:	reload	"	"	"	
	reduce stoppage	"	"	"	
3. Operate weapon:	soldier-weapon interface	"	"	"	
	marksmanship factors	"	"	"	
II. Engage Targets Individually					
4. Assault mode:	aimed fire	No. Hits/No. Rounds Fired			
5. Defensive position:	aimed fire	"	"		
6. Patrol operation:	aimed fire	"	"		

Following the specification of OPMs, the next requirement is to identify relevant PCVs (step 2.5.0). In the case of the M16A1 D-PAC evaluation, relevant PCVs are listed as follows:

1. Multiple Targets
2. Friendly and Hostile Targets
3. Variable Range
4. Target Movement:
   a. Direction
   b. Distance
   c. Rate
5. Target Exposure:
   a. Amount
   b. Duration
   c. Frequency
6. Target Camouflage
7. Target Termination When Hit
8. Target-Background Contrast Ratio
9. Terrain Features
10. Target Illumination

Step 2.6.0 involves the application of the five prioritization factors to the performances under consideration. This step requires the user first to rank the five factors in order of importance. For the current example, these rankings are given as follows:

1. Consequences of Inadequate Performance (CIP)
2. Task Importance (TI)
3. Frequency of Performance (FP)
4. Time Delay Tolerance (TDT)
5. Task Decay Rate (TDR)

Ratings on the prioritization factors are provided in the following table.

Performance			Factor				
			CIP	TI	FP	TDR	TDR
I. Prepare to Engage							
	1. Condition to operate:	calibrate weapon	M	M	L	M	L
		load weapon	H	M	H	H	L
	2. Maintain operation:	reload	H	M	H	H	L
		reduce stoppage	H	M	H	H	L
	3. Operate weapon:	soldier-weapon interface	M	M	H	H	M
		marksmanship factors	M	M	H	H	M
II. Engage Targets Individually							
	4. Assault mode:	aimed fire	H	M	L	H	M
	5. Defensive position:	aimed fire	H	M	L	H	M
	6. Patrol operation:	aimed fire	H	M	L	H	M

Using the information given above, performance priority rankings were obtained and are listed as follows:

1. Assault mode: aimed fire
2. Defensive Position: aimed fire
3. Patrol operation: aimed fire
4. Load weapon
5. Reload weapon
6. Reduce stoppage
7. Soldier-weapon interface
8. Marksmanship factors
9. Calibrate weapon

(Items 1–3: Tie; items 4–6: Tie; items 7–8: Tie)

The last step in phase 2 (step 2.7.0) involves deriving utility scores for the performances under consideration. Following the rating procedure outlined in Appendix E, utility scores were obtained and are presented in Table 4-2.

Table 4-2. Utility Matrix

Performance			Utility Score RE	Utility Score TM
I. Prepare to Engage				
	1. Condition to operate:	calibrate weapon	.00	.06
		load weapon	.00	.02
	2. Maintain operation:	reload	.03	.08
		reduce stoppage	.03	.24
	3. Operate weapon:	soldier-weapon interface	.03	.06
		marksmanship factors	.10	.34
II. Engage Targets Individually				
	4. Assault mode:	aimed fire	.13	.02
	5. Defensive position:	aimed fire	.57	.13
	6. Patrol operation:	aimed fire	.10	.05

The order of the utility matrix, denoted U, is $\underline{j}$ performances by $\underline{k}$ WDs.

Concept Definition

Phase 3 of the CIEA process is concerned with the definition of alternative D-PAC concepts. For fielded systems, this aspect of the procedure consists of integrating one or more training devices, or performance evaluation vehicles, into a set of D-PAC alternatives for analysis. In the case of emerging materiel systems, the utilities data produced in step 2.7.0 can be used as a guide to the specification of a series of conceptual D-PAC alternatives. These alternatives must, of course, be developed within the framework of the projected training device system. Whether based upon existing or projected devices, the resulting D-PAC alternatives must be specified at a level of detail that will support the requirements of the remaining aspects of the CIEA and permit reasonably precise LCCEs to be developed.

To illustrate how a D-PAC alternative might be generated from a set of devices, consider again the exemplary M16A1 analysis. Two devices, RF and WP, have been selected for study. The current method used to assess marksmanship proficiency is RF conducted one time per year. During RF, each soldier is taken to a firing range and assessed in a 40-round live-fire exercise. Prone and foxhole firing positions are employed; range and target exposure time also vary. A candidate is rated as "qualified" if he/she achieves 17 (23 at Ft. Benning) or more hits out of 40 possible. Figure 4-3 presents the firing positions, target ranges, and times currently used in RF.

WP is an M16A1 remedial marksmanship trainer designed to isolate individual performance deficiencies. A simulated M16A1 rifle is equipped with a target sensor and each target contains a light emitting diode (LED) which is sensed by the target sensor on the rifle. A predicted round impact point is determined by the LED-target sensor alinement. WP has a memory for recording up to 32 predicted shot impacts and a printer for providing a

TABLE 1 FOXHOLE POSITION

RD	Range (M)	Time (Sec)	Hit	Miss	No Fire
1	50	5			
2	200	5			
3	100	5			
4	150	5			
5	300	10			
6	250	10			
7	50	5			
8	200	5			
9	150	5			
10	250	10			
		TOTAL			

TABLE 2 PRONE POSITION

RD	Range (M)	Time (Sec)	Hit	Miss	No Fire
1	100	15			
2	250				
3	300	10			
4	50	10			
5	200				
6	150	15			
7	300				
8	50	10			
9	200				
10	100	5			
		TOTAL			

TABLE 3 PRONE POSITION

RD	Range (M)	Time (Sec)	Hit	Miss	No Fire
1	150	5			
2	100	10			
3	100	10			
4	200				
5	150	15			
6	250				
7	100	15			
8	300				
9	200	15			
10	100				
		TOTAL			

TABLE 4 FOXHOLE POSITION

RD	Range (M)	Time (Sec)	Hit	Miss	No Fire
1	100	10			
2	200				
3	250	20			
4	300				
5	100	15			
6	250				
7	250	10			
8	300	10			
9	50	10			
10	100				
		TOTAL			

Figure 4-3. Record Fire Scoring Table

printout of all shots on selected targets. Rifle recoil is simulated with recoil energy being variable from no recoil to a recoil intensity 40 perccent greater than the recoil of a standard M16A1 rifle. Three types of magazines are provided for use with the rifle: a 20-round (unlimited fire) magazine, a 30-round (unlimited fire) magazine, and a limited fire magazine that allows from 1 to 30 simulated rounds in the magazine. A headset is provided for simulating the firing sound of an M16A1 rifle. The WP also includes a selection for random misfire.

WP can present three targets: a scaled 25 meter zeroing target, a scaled 100 meter 'E' type silhouette target, and a 250 meter 'E' type silhouette target. Any target selected can be raised at random during a 1 to 9 second timeframe and can remain in a raised position for a duration of 2, 3, 5, 10, 15, 20, 25 seconds, or continuous. The WP provides a target 'Kill' function: a selection that will cause a raised target to drop when it is hit. Firing pads used with the WP provide the capability for the firer to fire from the foxhole or prone position.

A video display allows an observer to monitor individual shots and replay the last 3 seconds of each of the first 3 shots. Scoring available with the WP video display includes: the target on display, the number of hits on the target, the number of misses, late shots (fired after target drops), the shot number, and the total number of shots fired (Spartanics, Inc., 1976).

The individual devices were integrated with an evaluation scenario to form five D-PAC alternatives. These alternatives are listed as follows:

1. RF one time per year [RF(1)] (baseline).
2. RF twice per year [RF(2)].
3. RF four times per year [RF(4)].
4. RF once plus WP once [RF + WP].
5. RF once plus WP three times [RF + WP(3)].

The use of WP alone was ruled out in advance as being unacceptable.

To complete the specification of D-PAC alternatives, Table 4-3 presents the performance measurement methods for each constituent device.†

†In many applications, the consideration of measurement methods is not done until device capabilities have been characterized; i.e., after step 4.1.0.

Table 4-3

Performance Measurement Methods for Exemplary M16A1 D-PAC Devices

Performance			Device: RF	Device: WP
I. Prepare to Engage				
1.	Condition to operate:	calibrate weapon	observor by verification (OBV)	OBV; machine-assisted
		load weapon	OBV of SOP	OBV of SOP
2.	Maintain operation:	reload	OBV of SOP	OBV of SOP
		reduce stoppage	OBV of SOP	OBV of SOP
3.	Operate weapon:	soldier-weapon interface	OBV	OBV; machine assisted
		marksmanship factors	OBV	OBV; machine-assisted
II. Engage Targets Individually				
4.	Assault mode:	aimed fire	Observor count (OC)	Machine Count (MC) with hard copy printout (PO)
5.	Defensive position:	aimed fire	OC	MC w/PO
6.	Patrol operation:	aimed fire	OC	MC w/PO

Concept Evaluation

Once alternative D-PAC concepts have been defined, the stage is set for phase 4, Concept Evaluation. The objective of phase 4 is, first, to characterize each D-PAC alternative in terms of its effectiveness, or benefits. In this context, D-PAC effectiveness is defined as the extent to which an alternative provides complete and timely information on all performances relevant to the proposed D-PAC application. The D-PAC alternatives are then subjected to a trade-off analysis in which information benefit is weighted against the costs associated with the various effectiveness-producing features.

The general philosophy of the D-PAC effectiveness evaluation scheme is depicted in Figure 4-4. Following Figure 4-4, three general attributes of D-PAC alternatives combine to produce effectiveness. These attributes are: (1) device capabilities, (2) performance measurement system characteristics, and (3) the application scenario (i.e., the frequency with which performance status information is obtained). Changes in any of the three characteristics can change the worth of a D-PAC alternative in application. Those portions of the CIEA methodology directed at characterizing D-PAC alternatives along the three primary effectiveness components and then trading off the result against system costs is described in the following paragraphs.

Obtain Device Capabilities Matrix

The first step in phase 4 involves characterizing each device in terms of its potential for performance assessment. This step is carried out by developing a Device Capabilities matrix, denoted C, of order $\underline{k}$ performances by $\underline{n}$ devices. Entries in C are either "1" or "0", depending upon whether devices do or do not provide a vehicle for the evaluation of specific performances.

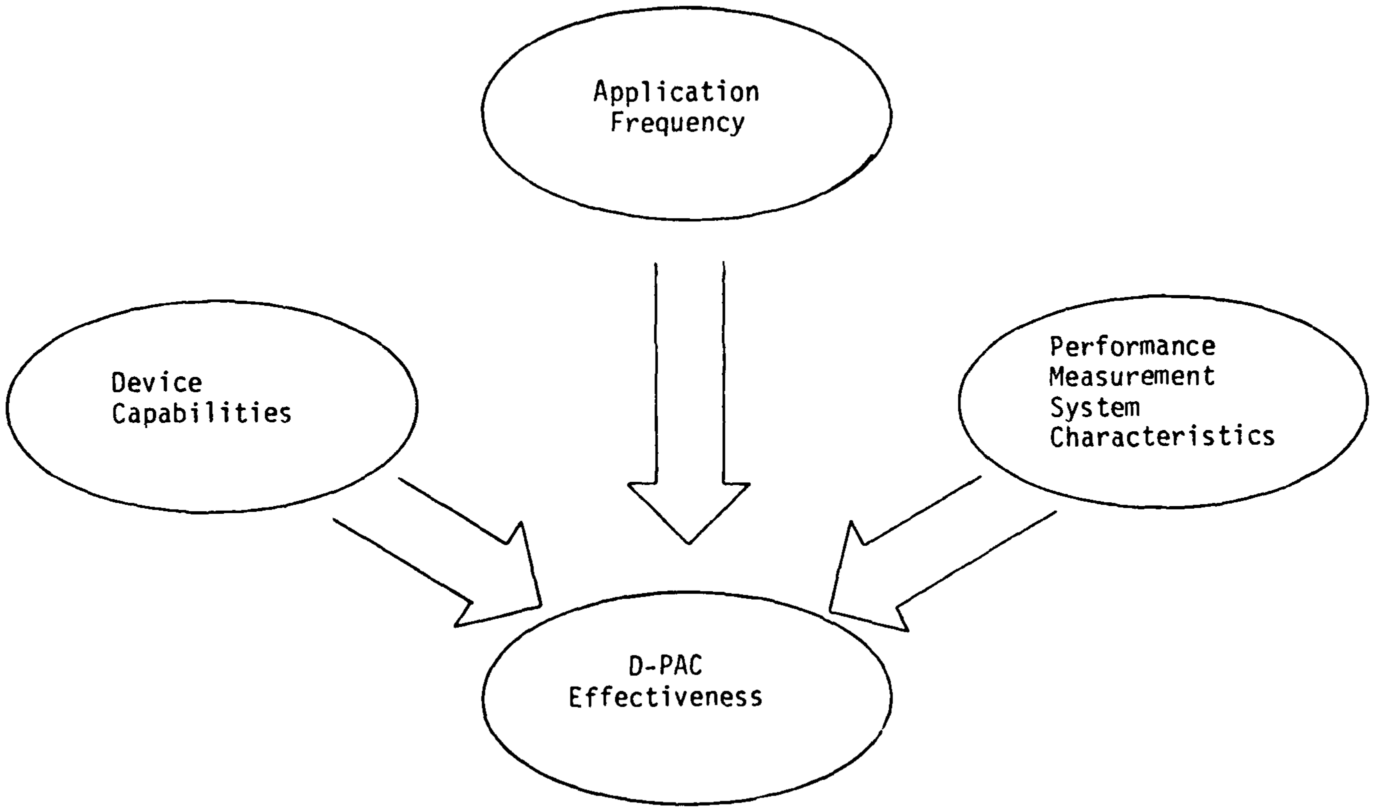


Figure 4-4. Concept of D-PAC Effectiveness Evaluation

To illustrate the structure of a Device Capabilities matrix, the C matrix for the M16A1 example is presented as Table 4-4.

Table 4-4

Device Capabilities Matrix

Performance		Record Fire (RF)	Weaponeer (WP)
I. Prepare to Engage			
1. Condition to operate:	calibrate weapon	1	1
	load weapon	1	1
2. Maintain operation:	reload	1	1
	reduce stoppage	1	1
3. Operate weapon:	soldier-weapon interface	1	1
	marksmanship factors	1	1
II. Engage Targets Individually			
4. Assault mode:	aimed fire	1	1
5. Defensive position:	aimed fire	1	1
6. Patrol operation:	aimed fire	1	1

Note that in the exemplary exercise both devices address all performances of interest.

Obtain Measurement Precision Ratings

In the next step (4.2.0), each cell of the C matrix containing a "1" (i.e., performance assessment is possible) is elaborated upon by obtaining precision ratings for the methods used to obtain performance status information (see, for example, Table 4-3). Using the rating procedure described in Appendix F, SMEs provide MP ratings for each device on each performance.

The range for individual MP scores is from 0 to 100. In assigning ratings, SMEs are asked to consider both content validity (i.e., the comprehensiveness of the OPM) and the reliability (i.e., stability upon replication) of the measurement procedure. The order of the MP matrix is k performances by n devices.

As an example, MP scores for the M16A1 CIEA are shown in Table 4-5.

Table 4-5

Measurement Precision Matrix

Performance			MP Ratings RF	MP Ratings WP
I. Prepare to Engage				
1.	Condition to operate:	calibrate weapon	75	50
		load weapon	90	10
2.	Maintain operation:	reload	90	10
		reduce stoppage	80	5
3.	Operate weapon:	soldier-weapon interface	50	85
		marksmanship factors	70	80
II. Engage Targets Individually				
4.	Assault mode:	aimed fire	65	80
5.	Defensive position:	aimed fire	65	80
6.	Patrol operation:	aimed fire	65	80

Obtain Performance Context Matrix

The objective of step 4.3.0 is to characterize each component device in terms of its coverage of PCVs. Recall that PCVs are condition variables, target characteristics, and the like that can be significant performance moderators in an operational environment. The explicit treatment of PCVs represents an attempt to amplify the results of step 4.1.0 by obtaining quantitative indices of the degree to which each device provides a vehicle for realistic performance assessment under conditions likely to be encountered in an operational environment. The conduct of step 4.3.0 is a seven substep process, with each of the substeps described as follows.

<u>Obtain Context Variable Importance Vector</u>. In the first substep, each PCV judged relevant to the performances under consideration (identified in step 2.5.0) is assigned a weight reflecting its relative importance for consideration in the projected D-PAC. The weighting procedure used in this substep is the same as that used to weight WDs (described in Appendix C). To illustrate the results of substep 4.3.1, the PCV importance vector, denoted <u>I</u>, for the M16A1 demonstration CIEA is presented as Table 4-6. Note that the elements of the vector <u>I</u> are normalized to sum to 100.

Table 4-6

PCV Importance Vector

Condition Variable	Weight Vector, I
1. Multiple Targets	13
2. Friendly and Hostile Targets	3
3. Variable Range	20
4. Target Movement:	14
a. Direction	(8)
b. Distance	(2)
c. Rate	(4)
5. Target Exposure:	8
a. Amount	(4)
b. Duration	(3)
c. Frequency	(1)
6. Target Camouflage	0
7. Target Termination When Hit	19
8. Target-Background Contrast Ratio	6
9. Terrain Features	7
10. Target Illumination	2

Obtain Device Coverage Incidence Matrix. Substep 4.3.2 involves characterizing each device according to its coverage of designated PCVs. Accordingly, a "0", "1" characterization scheme is again used to this end. A value of "1" indicates that the device adequately addresses the PCV. A "0" score indicates that the requirements of the PCV are not met by the device.

To illustrate the rating process, Table 4-7 presents the Device Coverage Incidence matrix (denoted X) for the M16A1 example. Again, an entry of "1" indicates that the device addresses a particular PCV; an entry of "0" indicates that the device does not include the capability.

Table 4-7

Device Coverage Incidence Matrix

Context Variable	Device Coverage RF	Device Coverage WP
1. Multiple Targets	1	1
2. Friendly and Hostile Targets	0	1
3. Variable Range	1	1
4. Target Movement:		
a. Direction	0	0
b. Distance	0	0
c. Rate	0	0
5. Target Exposure:		
a. Amount	0	0
b. Duration	1	1
c. Frequency	1	1
6. Target Camouflage	0	0
7. Target Termination When Hit	1	1
8. Target-Background Contrast Ratio	0	0
9. Terrain Features	0	0
10. Target Illumination	1	1

The order of the matrix X is m context variables by n devices.

Obtain Absolute Coverage Matrix. In substep 4.3.3, the PCV importance vector I is multiplied, element by element, by the Device Coverage Incidence matrix X to obtain the Absolute Coverage matrix, denoted A. The matrix A augments the results of substep 4.3.2 by transferring a "relative value" index into each of the locations of the X matrix that contain a "1". Since it is an augmentation of X, the matrix A is also of order m condition variables by n devices. As an example, the A matrix for the M16A1 CIEA is provided in Table 4-8.

Table 4-8

Absolute Coverage Matrix

Condition Variable	Coverage X Importance	
	RF	WP
1. Multiple Targets	13	13
2. Friendly and Hostile Targets	0	3
3. Variable Range	20	20
4. Target Movement:		
a. Direction	0	0
b. Distance	0	0
c. Rate	0	0
5. Target Exposure:		
a. Amount	0	0
b. Duration	3	3
c. Frequency	1	1
6. Target Camouflage	0	0
7. Target Termination When Hit	19	19
8. Target-Background Contrast Ratio	0	0
9. Terrain Features	0	0
10. Target Illumination	2	2

Obtain Performance Relevancy Matrix. Following the derivation of the Absolute Coverage matrix, substep 4.3.4 concerns the construction of the Performance Relevancy matrix, denoted R. The issue of performance relevancy is addressed because it is recognized that not all PCVs are important, or relevant, in the assessment of all performances. The Performance Relevancy matrix is an incidence matrix of order $\underline{k}$ performances by $\underline{m}$ context variables. Entries in R are either "1" or "0" indicating, respectively, that specific PCVs are or are not relevant to the assessment of given performances. As an example of a representative Performance Relevancy matrix, the R matrix for the M16A1 performances and context variables is presented as Table 4-9.

Compute Normalization Constants. The next substep in phase 4 is to obtain normalization constants, denoted n_k. For each performance, a scalar normalization constant is obtained by forming the vector product

$$n_k = \underline{r}_k \underline{I} = \sum_m r_{km} i_m . \qquad (4\text{-}1)$$

In (4-1), $\underline{r}_k$ is the k^{th} row of the Performance Relevancy matrix R, and $\underline{I}$ is the PCV importance vector; or, in algebraic notation,

r_{km} is the entry in the k^{th} row and m^{th} column of R

and i_m is the m^{th} entry in the importance vector $\underline{I}$.

The n_k reflect the weighted context variable coverage relevant to each performance. As an example, the normalization constants for the M16A1 example are given in the last column of the Relevancy matrix (Table 4-9).

Obtain Relevant Coverage Matrix. In substep 4.3.6, the Absolute Coverage matrix, A, is screened by the Relevant Capabilities matrix, R, to form the Relevant Coverage matrix, denoted RCM. RCM is obtained by forming the matrix product

$$RCM = R\ A \qquad (4\text{-}2)$$

Table 4-9. Exemplary Performance Relevancy Matrix

Performance \ Context Variable			1. Multiple Targets	2. Friendly and Hostile Targets	3. Variable Range	4. Target Movement: a. Direction	b. Distance	c. Rate	5. Target Exposure: a. Amount	b. Duration	c. Frequency	6. Target Camouflage	7. Target Termination When Hit	8. Target-Background Contrast Ratio	9. Terrain Features	10. Target Illumination	Normalization Constant, n_k
I. Prepare to Engage																	
1.	Condition to Operate:	calibrate weapon	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		load weapon	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2.	Maintain Operation:	reload	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		reduce stoppage	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.	Operate Weapon:	soldier-weapon interface	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		marksmanship factors	1	0	0	1	1	1	1	1	1	0	0	0	0	0	35
II. Engage Targets Individually																	
4.	Assault Mode:	aimed fire	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100
5.	Defensive Position:	aimed fire	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100
6.	Patrol Operation:	aimed fire	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100

The order of RCM is k performances by n devices. Entries in RCM reflect weighted PCV coverage by devices. For example, the RCM matrix for the M16A1 CIEA is given as Table 4-10. The entry "58" for the performance "Patrol Operation: aimed fire" under RF indicates that the relative importance ratings for the PCVs covered by RF, and relevant to the indicated performance, sum to 58.

Table 4-10

Relevant Coverage Matrix

Performance			Device RF	Device WP
I. Prepare to Engage				
	1. Condition to operate:	calibrate weapon	0	0
		load weapon	0	0
	2. Maintain operation:	reload	0	0
		reduce stoppage	0	0
	3. Operate weapon:	soldier-weapon interface	0	0
		marksmanship factors	17	17
II. Engage Targets Individually				
	4. Assault mode:	aimed fire	58	61
	5. Defensive position:	aimed fire	58	61
	6. Patrol operation:	aimed fire	58	61

Obtain Performance Context Matrix. The final substep in step 4.3.0 involves normalizing the entries in RCM to reflect the weighted proportion of context variable coverage that is actually relevant to specific performances. This substep is carried out by dividing the entries in RCM by the appropriate normalization constant:

$$pc_{kn} = \frac{rc_{kn}}{n_k} \times 100, \tag{4-3}$$

where pc_{kn} is the performance context (PC) "coverage" rating for the k^{th} performance on the n^{th} device;

n_k is the normalization constant for the k^{th} performance (obtained in substep 4.3.5);

and rc_{kn} is the relative coverage rating for the k^{th} performance on the n^{th} device (from substep 4.3.6).

The result is multiplied by 100 to scale the result to fall between zero and 100. In the event that one of the normalization constants is zero, indicating that no PCVs are relevant to the assessment of that performance, the resulting entry in the PC matrix (which will be "0/0") is arbitrarily defined to be 100.

The PC matrix (of order k performances by n devices) for the M16A1 example is given as Table 4-11.

Table 4-11

Exemplary Performance Context Matrix

Performance		Device RF	Device WP
I. Prepare to Engage			
1. Condition to operate:	calibrate weapon	100	100
	load weapon	100	100
2. Maintain operation:	reload	100	100
	reduce stoppage	100	100
3. Operate weapon:	soldier-weapon interface	100	100
	marksmanship factors	49	49
II. Engage Targets Individually			
4. Assault mode:	aimed fire	58	61
5. Defensive position:	aimed fire	58	61
6. Patrol operation:	aimed fire	58	61

Compute Device Measurement Effectiveness Matrix

The stage is now set for integrating the MP ratings (obtained in step 4.2.0) and the Performance Context ratings (the results of step 4.3.0) to form Device Measurement Effectiveness (DME) scores. DME scores are computed in the following manner:

$$dme_{kn} = \frac{mp_{kn}\ pc_{kn}}{10^2}, \qquad (4\text{-}4)$$

where dme_{kn} represents the DME score for the n^{th} device on the k^{th} performance,

mp_{kn} is the MP rating of the n^{th} device on the k^{th} performance,

pc_{kn} is the PC score for the n^{th} device on the k^{th} performance,

and 10^2 is a scaling constant.

In matrix notation, (4-4) is expressed as

$$DME = \left(\frac{1}{10^2}\right) MP * PC, \qquad (4\text{-}5)$$

where the symbol "*" denotes element-by-element matrix multiplication.

To illustrate the derivation of DME scores, Table 4-12 presents the values for the M16A1 D-PAC evaluation.

Table 4-12

Device Measurement Effectiveness Matrix

Performance			Device Measurement Effectiveness RF	WP
I. Prepare to Engage				
	1. Condition to operate:	calibrate weapon	75	50
		load weapon	90	10
	2. Maintain operation:	reload	90	10
		reduce stoppage	80	5
	3. Operate weapon:	soldier-weapon interface	50	85
		marksmanship factors	34	39
II. Engage Targets Individually				
	4. Assault mode:	aimed fire	38	49
	5. Defensive position:	aimed fire	38	49
	6. Patrol operation:	aimed fire	38	49

Obtain Alternative Measurement Effectiveness Matrix

The next step in the improved CIEA procedure is to combine DME ratings across devices to obtain effectiveness ratings for D-PAC alternatives. Measurement effectiveness ratings for D-PAC alternatives are obtained by computing the weighted mean of the DME ratings for their component devices; that is,

$$ame_{ki} = \frac{\sum_n dme_{kn} f_{ni}}{\sum_n f_{ni}} \quad . \tag{4-6}$$

In (4-6), ame_{ki} represents the alternative measurement effectiveness (AME) score for the i^{th} D-PAC alternative on the k^{th} performance,

dme_{kn} is the DME score for the n^{th} device on the k^{th} performance,

and f_{ni} is the frequency with which the n^{th} device is used in the i^{th} alternative. For example, the fifth M16A1 D-PAC alternative specifies the use of RF once and WP three times per year. In this case $f_{RF,5} = 1$ and $f_{WP,5} = 3$.

The AME matrix for the M16A1 example is provided as Table 4-13.

Obtain Frequency Utility Ratings for Performance Domains

In step 4.6.0, SMEs are asked to provide frequency utility (FU) ratings for each of the performance domains (i.e., functional groupings for performances) under consideration. Using the guidelines presented in Appendix G, FU ratings are provided on a 0-to-100 scale. The scores reflect the utility of receiving status information on the performances nested under each domain at the frequency indicated (i.e., 1, 2, 3, or 4 times per year). If desired, FU ratings can be obtained separately for each WD. However, in the M16A1 example to follow, only one set of FU ratings is provided. These exemplary ratings are provided in Table 4-14.

Table 4-14

Frequency Utility Matrix

		Evaluation Frequency			
		1	2	3	4
Performance Domain	I.	50	60	78	80
	II.	60	70	80	90

Table 4-13. AME Matrix for M16A1 D-PAC Evaluation

Performance		Alternative Measurement Effectiveness RF(1) (1)	RF(2) (2)	RF(4) (3)	[RF+WP] (4)	[RF+WP(3)] (5)
I. Prepare to Engage						
1. Condition to Operate:	calibrate weapon	75	75	75	62.5	56.25
	load weapon	90	90	90	50	30
2. Maintain Operation:	reload	90	90	90	50	30
	reduce stoppage	80	80	80	42.5	23.75
3. Operate Weapon:	soldier-weapon interface	50	50	50	67.5	76.25
	marksmanship factors	34	34	34	36.5	37.75
II. Engage Targets Individually						
4. Assault Mode:	aimed fire	38	38	38	43.5	46.25
5. Defensive Position:	aimed fire	38	38	38	43.5	46.25
6. Patrol Operation:	aimed fire	38	38	38	43.5	46.25

Compute Alternative Effectiveness Matrix

The next step in the analysis involves integrating AME scores, which reflect device capabilities and the precision of the performance assessment system, with FU ratings to obtain an Alternative Effectiveness (AE) score for each D-PAC option on each performance. AE ratings are computed in the following manner:

$$AE_{ji(j)} = \frac{ame_{ki} \; fu_{ki(j)}}{10^2} , \tag{4-7}$$

where $ae_{ki(j)}$ is the AE score for the i^{th} D-PAC alternative on the k^{th} performance nested under the j^{th} WD (if applicable),

ame_{ki} is the measurement effectiveness score of the i^{th} alternative for the k^{th} performance,

$fu_{ki(j)}$ is the FU ratings of the i^{th} alternative on the k^{th} performance (nested under a particular performance domain) for the j^{th} WD (if applicable),

and 10^2 is a scaling constant.

For the M16A1 example, AE ratings are presented in Table 4-15.

Compute Partial Information Utility Matrix

As a next step in the analysis, AE ratings are combined across performances to obtain Partial Information Utility (PIU) scores for each alternative on each WD. Entries in the PIU matrix (of order j WDs by i alternatives) are obtained using the following combination rule:

$$piu_{ji} = \sum_k u_{jk} \; ae_{ki(j)} . \tag{4-8}$$

Table 4-15. Alternative Effectiveness Ratings

Performance		Alternative (1)	(2)	(3)	(4)	(5)
I. Prepare to Engage						
1. Condition to Operate:	calibrate weapon	33.75	45	60	37.5	45
	load weapon	45	54	72	30	24
2. Maintain Operation:	reload	45	54	72	30	24
	reduce stoppage	40	48	64	25.5	19
3. Operate Weapon:	soldier-weapon interface	25	30	40	40.5	61
	marksmanship factors	17	20.4	27.2	21.9	30.2
II. Engage Targets Individually						
4. Assault Mode:	aimed fire	22.8	26.6	30.4	30.45	41.63
5. Defensive Position:	aimed fire	22.8	26.6	30.4	30.45	41.63
6. Patrol Operation:	aimed fire	22.8	26.6	30.4	30.45	41.63

In (4-8), piu_{ji} denotes the PIU score for the i^{th} alternative on the j^{th} WD,

u_{jk} is the utility score for the k^{th} performance nested under the j^{th} WD (obtained in step 2.7.0),

and $ae_{ki(j)}$ is the AE score of the i^{th} alternative on the k^{th} performance nested under the j^{th} WD (if applicable).

In matrix notation, an expression equivalent to (4-8) is

$$PIU = U\, AE_j \,, \tag{4-9}$$

where PIU is the matrix of PIU scores (j x i),

U is the utilities matrix (j x k),

and AE_j is the matrix of AE scores (k x i) for the j^{th} WD.

The matrix of PIU scores for the M16A1 example is provided as Table 4-16.

Table 4-16

Partial Information Utility Matrix

		D-PAC Alternative				
		(1) RF(1)	(2) RF(2)	(3) RF(4)	(4) [RF+WP]	(5) [RF+WP(3)]
WD	RE	23.25	26.08	32.32	29.43	39.44
	TM	27.96	33.67	43.88	27.33	31.91

Compute Information Utility Scores for Alternatives

PIU scores are next combined across WDs to obtain a global Information Utility (IU) score for each D-PAC alternative. In CIEA, the IU scores represent the aggregate measure of benefit for D-PAC alternatives. The combination rule for aggregating PIU scores is given as follows:

$$IU_i = \sum_j W_j \, piu_{ji} \, , \tag{4-10}$$

where IU_i is the IU score for the i^{th} alternative,

W_j is the importance weight of the j^{th} WD (obtained in step 2.1.0),

and piu_{ji} is the PIU score for the i^{th} alternative on the j^{th} WD.

In matrix notation (4-10) is given as the vector-matrix product

$$\underline{IU}_i' = \underline{W}_j' \, PIU \, . \tag{4-11}$$

Continuing with the M16A1 example, the vector of IU scores is given in Table 4-17.

Table 4-17

M16A1 D-PAC IU Vector

	D-PAC Alternative				
	(1)	(2)	(3)	(4)	(5)
IU	26.41	31.16	40.06	28.02	34.39

Estimate Life-Cycle Costs of Alternatives

Phase 4 continues with the development of LCCEs for D-PAC alternatives. Although cost estimation is formally considered at this point in the procedure, the cost analysis actually may be initiated any time after D-PAC alternatives have been defined (i.e., following phase 3, Concept Definition). The cost analysis should, in fact, be initiated as early as possible, since this aspect of CIEA will usually prove to be somewhat time-consuming.

The objective of step 4.10.0 is to produce an LCCE for each D-PAC alternative; that is, to provide an estimate of what each alternative will cost to develop and deploy and then to operate and maintain over its projected service life. To assist in the development of cost estimates for alternatives, Hawley and Dawdy (1981a) present a structured D-PAC costing guide. The guide leads an analyst through the steps of a D-PAC cost analysis beginning with a determination of the anticipated facility load and ending with a total estimated cost for each alternative over its service life. It should be noted that the cost estimates provided by the costing guide consider only those design, development (e.g., testing to validate measures and establish standards), and administration (e.g., testing, data processing) costs which would be incurred over and above those associated with design, development, and use of the devices for training.

Summarize Results in Alternative-Versus-Criteria Array

Following the determination of LCCEs for D-PAC alternatives, the next step in phase 4 involves a summarization of the results of the analysis in the Alternative-Versus-Criteria array. This array displays PIU scores, IU scores, LCCEs, and Relative Information Cost (RIC) scores by D-PAC alternatives. RIC scores are obtained by dividing the LCCE for each alternative by that of the option designated as baseline (i.e., either the present capability or the most conventional D-PAC alternative); that is,

$$RIC_i = LCC_i/LCC_b \; . \tag{4-12}$$

In most situations, the selection of a preferred D-PAC alternative will be made on the basis of the entries provided in the Alternative-Versus-Criteria array.

To illustrate the form of the Alternative-Versus-Criteria array, again consider the hypothetical analysis of a set of D-PAC alternatives for the M16A1 rifle. Assume that cost estimates for the five alternatives have been determined as follows; for reader convenience, RIC figures are also provided.

Alternative	LCCE($000's)	RIC
1. RF(1)	261.8	1.00
2. RF(2)	467.4	1.79
3. RF(4)	846.0	3.23
4. RF+WP	341.7	1.31
5. RF+WP(3)	438.2	1.67

The complete Alternative-Versus-Criteria array is then assembled as follows.

Table 4-18

Alternative-Versus-Criteria Array

	Criterion				
	PIU				
Alternative	RE	TM	IU	LCC	RIC
1. RF(1)	23.25	27.96	26.41	261.8	1.00
2. RF(2)	26.08	33.67	31.16	467.4	1.79
3. RF(4)	32.32	43.88	40.06	846.0	3.23
4. RF+WP	29.43	27.33	28.02	341.7	1.31
5. RF+WP(3)	39.44	31.91	34.39	438.2	1.67

Determine Most Cost and Information Effective Alternative

The final step in phase 4 concerns selecting a preferred D-PAC alternative from among those under consideration. Recall from the summary material presented in section 1 that the preliminary CIEA methodology prescribed the use of Relative Information Worth (RIW) (i.e., RIU/RIC) as a criterion for the selection of a preferred D-PAC alternative. One caveat in this approach is that the use of RIW is predicated upon the tenability of the equal-interval scaling assumption for IU. It should be noted, however, that the results presented in section 3 of this report cast doubt upon the tenability of the equal-interval assumption. As noted therein, the equal-interval assumption may generally hold for a particular group, but even then it may not be tenable across the entire scale range for entities being rated. That is, IU may not be uniformly equal-interval, even for a carefully selected and well trained user group. The implications of violations of the equal-interval assumption are significant in the identification of a preferred D-PAC alternative. In essence, they mean that the consideration of RIW is not warranted in most instances and may, in fact, be misleading.

In view of the results presented in section 3, a more reasonable approach to identifying a preferred alternative is what might be termed a "benefit-affordability" strategy. That is, the objective of step 4.12.0 should be the identification of the most information effective D-PAC from among those alternatives that are judged to be affordable. Such an analysis can be carried out within a set partitioning framework similar to the simple branch and bound procedures used in operations research (for example, see Hillier & Lieberman, 1980).

The first substep in a benefit-affordability approach to step 4.12.0 is to partition the D-PAC alternatives into two sets--designated as "acceptable" and "unacceptable"--on the basis of their IU ratings. For example, all D-PAC alternatives that rate lower than the baseline case on IU could be classified as unacceptable, while all other alternatives are classified as acceptable. Unacceptable alternatives are then eliminated from further consideration.

In substep two, those alternatives judged acceptable on the IU criterion are next evaluated in terms of their LCCE. Again, two classes of alternatives are designated: acceptable (affordable) and unacceptable (not affordable). The top two, or possibly three, alternatives in the IU acceptable-LCCE acceptable class are then subjected to additional scrutiny in what might be termed a quasi cost-effective analysis. First, the remaining alternatives are ranked on IU. Next, LCCEs are listed. If the top-ranked alternative (on IU) is also the lowest cost option, then the choice is simple: select that alternative as preferred. If, on the other hand, the top-ranked candidate is not the least costly (probably a more usual situation), it is then necessary to judge whether the increased utility, or benefit, accruing from the top-ranked choice is worth its additional cost; that is, to decide whether or not a fair trade-off between incremental benefit and cost would be made in selecting the top-ranked alternative. If the result of this judgment is "no", then the second-rated choice is preferred. This procedure, or a similar method, can be repeated for any number of alternatives remaining after the IU and LCCE partitionings have been carried out.

To illustrate the benefit-affordability approach to selecting a preferred D-PAC alternative, again consider the M16A1 example. The first criterion, IU, does not remove any of the alternatives from consideration. The baseline case of RF(1) has the lowest IU score. Hence, at substep two the following alternatives are considered on the basis of their cost:

Alternative	IU	LCC	RIC
RF(4)	40.06	846.0	3.23
RF+WP(3)	34.39	438.2	1.67
RF(2)	31.16	467.4	1.79
RF+WP	28.02	341.7	1.31
RF(1)	26.41	261.8	1.00

In substep two, a preliminary decision is made that RF(4) is too costly (i.e., not affordable), thus it is eliminated from further consideration. Four choices, listed as follows, now remain:

Alternative	IU	ΔIU%	LCC	RIC	ΔC%
RF+WP(3)	34.39	10.4	438.2	1.67	-6.7
RF(2)	31.16	11.2	467.4	1.79	36.6
RF+WP	28.02	6.1	341.7	1.31	31.0
RF(1)	26.41	---	261.8	1.00	---

The alternative designated RF(2) also is an obvious choice for elimination on the basis of cost since it has a higher LCCE than the now top-rated alternative, RF+WP(3). Once RF(2) is eliminated, three choices remain in the analysis. These choices are summarized as follows.

Alternative	IU	ΔIU%	LCC	RIC	ΔC%
RF+WP(3)	34.39	22.7	438.2	1.67	27.5
RF+WP	28.02	6.1	341.7	1.31	31.0
RF(1)	26.41	---	261.8	1.00	---

A review of the above table indicates that the alternative RF+WP(3) results in a 22.7% increase in IU over RF+WP. The incremental benefit is obtained with a cost increment of 27.5 percent. Since this situation represents nearly a one-for-one benefit-cost trade-off, it is judged to be fair. Thus, alternative 5, RF+WP(3), is judged to be the preferred D-PAC option.

Admittedly, the benefit-affordability trade-off approach described herein is considerably more subjective than a straight cost-effectiveness strategy. As a result, different groups of decision-makers may select different D-PAC options as preferred. The benefit-affordability approach does, however, require decision-makers to consider both benefit and cost in selecting a preferred capability while avoiding the problems that make an analysis based upon the use of cost-benefit ratios somewhat hazardous.

Design Specifications

After identifying the preferred D-PAC alternative, the final phase in the CIEA process concerns the development of detailed design specifications. This final step is undertaken to provide design engineers with sufficient information to be able to develop a prototype of the desired capability. In effect, phase 5 serves as the bridge between the conceptualization and evaluation stages of D-PAC development and the construction and concept validation stages. Also as part of phase 5, plans for evaluating and validating the prototype capability are developed. For emerging systems, it is intended that the D-PAC be tested and evaluated along with the device(s) of which it is an integral part.

5. DISCUSSION

This report represents the first part of a two-volume series concerned with the development of a viable CIEA methodology. The material presented herein describes the structure for an improved CIEA procedure based upon the application of various MAUM procedures. Volume two of the series (Brett, Chapman, & Hawley, 1982) contains a detailed description of an application of the improved methodology to a series of D-PAC alternatives for gunnery training on the Combat Engineer Vehicle.

In addition to the work on methodological improvements, the current project also concerned the development of a computer-aided procedure for the conduct of CIEA ("Cost and Information Effectiveness Analysis: A Computer-Aided Approach," 1982). This computer program is currently implemented on an Apple II microcomputer. The program leads users through the analysis in a structured fashion. In this manner, confusion is avoided and users are relieved of the computational drudgery associated with a manual application of the methodology.

Whether the improved CIEA procedure is employed in a manual or computer-aided mode, results from the current year's effort suggest that several application guidelines be observed. The first guideline concerns the composition of the user group. Results from the tryout exercises conducted at Ft. Benning indicate that the procedure is very sensitive to the composition of the user group. This being the case, care should be taken to select users that are familiar with the subject materiel system and its training devices (existing or projected). In addition, users should be somewhat familiar with the objectives, rationale, and processes underlying the methodology. It is the authors' view that familiarity with the methodology obtained through its repeated application will result in improved results in terms of reliability and validity. In short, with CIEA, there is no substitute for application experience.

The second general guideline for the application of the improved methodology concerns the robustness of the analysis. Again, results from the current study indicate that the MAUM-based CIEA should be viewed as a decision aid rather than as a deterministic, or algorithmic, procedure. That is, users should not take the results of the analysis too literally. The complete exercise of the method requires users to consider the development of a D-PAC from several perspectives. However, the subjective nature of the MAUM procedures employed suggest that the resulting IU scores be reviewed critically. The movement away from the use of cost-effectiveness ratios and toward the benefit-affordability approach to system evaluation is a tacit recognition of this result.

In terms of future directions for CIEA refinement and development, one comment is in order. It is certain that sophisticated readers will find deficiencies in the current procedure. In all likelihood, these criticisms will involve a more thorough treatment of various aspects of D-PAC development and evaluation. The methods selected for use herein were selected on the basis of an extensive review of the MAUM/psychological scaling literature and the results of a series of formative tryouts. In addition, a large amount of considered judgment entered into the developmental process. It was necessary to trade-off subjectively, so to speak, the potential benefits of a more explicit handling of various aspects of the analysis against the liabilities accompanying the development of a more complex procedure. That being the case, perhaps the best way to proceed in further refining the CIEA methodology is to apply it across a range of situations and thereby to obtain additional information concerning the procedure's acceptability, its perceived validity, and the like. It may very well be, for example, that the present methodology is already too complex for the intended user population. It might thus prove beneficial to simplify the analysis rather than to increase its complexity still further. In the final analysis, the developmental effort described herein will have been successful if the CIEA methodology is actually employed in the development of D-PACs. Repeated applications will imply that users can indeed exercise the methodology and find the results of the analysis worth the effort required to obtain them.

References

American Airlines, Inc. Optimized flight crew training: A step toward safer operations. Ft. Worth, Tex.: American Airlines Flight Training Academy, April 1969.

Barish, N.N., & Kaplan, S. Economic analysis for engineering and managerial decision making (2d Ed.). New York: McGraw-Hill, 1978.

Bedford, N.M., & Onsi, M. Measuring the value of information--An information-theory approach. Management Services, January-February, 1966, 1, 15-22.

Bell, J.D., & Pickering, E.J. Use of performance measurement data from the 14A2 ASW team trainer complex in a performance assessment system. San Diego: Naval Personnel Research & Development Center, 1979. (AD A067-401)

Brett, B.E., Chapman, W.A., & Hawley, J.K. Cost and information effectiveness analysis: An improved methodology. Volume II. Valencia, PA: Applied Science Associates, Inc., 1982.

Callan, J.R., Kelley, R.T., & Nicotra, A. Measuring submaring approach officer performance on the 21A40 trainer: Instrumentation and preliminary results (NPRDC-TR-78-9). San Diego, Calif.: U.S. Naval Personnel Research and Development Center, January 1978.

Campbell, D.T., & Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Caro, P. Aircraft simulators and pilot training. Human Factors, 1973, 15(6), 502-509.

Cost and IEA: A Computer-Aided Approach. Valencia, PA: Applied Science Associates, Inc., 1982.

Crawford, A., & Brock, J. Using simulators for performance measurement. In Proceedings of the Symposium on Productivity Enhancement: Personnel Assessment in Navy Systems. San Diego, Calif.: Naval Personnel Research and Development Center, October 1977.

Department of Defense. Catalog of Logistics Models, 1980.

Dupuy, W.E., & Gorman, P.F. TRADOC mission and resources briefing. Ft. Monroe, Va.: U.S. Army Training & Doctrine Command, 1978. (videotape)

Finley, D.L., Gainer, C.A., & Muckler, F.A. Test and evaluation of a carrier air traffic control center team trainer as a performance qualification instrument: Phase I report. Canoga Park, Calif.: XYZYX Information Corp., 1974. (NTIS No. AD B003 213L)

Fishman, G.S. Principles of discrete event simulation. New York: Wiley, 1978.

Fitzpatrick, R., & MOrrison, E.J. Performance and product evaluation. In R.L. Thorndike (Ed.), Educational Measurement (2nd Ed). Washington, D.C.: American Council on Education, 1971.

Glaser, R., & Klaus, D.J. Proficiency measurement: Assessing human performance. In R.M. Gagne (Ed.) Psychological principles in system development. New York: Holt, Rinehart, and Winston, 1972.

Hawley, J.K., & Dawdy, E.D. Cost and information effectiveness analysis (CIEA): A methodology for evaluating a training device operational readiness assessment capability (DORAC). Valencia, Pa.: Applied Science Associates, Inc., February 1981. (a)

Hawley, J.K., & Dawdy, E.D. Training device operational readiness assessment capability (DORAC): Feasibility and utility. Valencia, Pa.: Applied Science Associates, Inc., April 1981. (b)

Hanson, D.N., Harris, D.A., & Ross, S. Flexilevel adaptive testing paradigm: Validation in technical training. (AFHRL-R-77-35(1)). Lowry AFB, Colorado: Technical Training Div., AFHRL, July 1977.

Hillier, F.S., & Lieberman, G.J. Introduction to Operations Research (3rd Ed.). San Francisco: Holden-Day, 1980.

Hopkins, C.O. How much should you pay for that box? In Proceedings of the 19th Annual Meeting of the Human Factors Society. Dallas, Tex.: Human Factors Society, 1975, i-vi.

Johnson, E.M., & Huber, P.G. The technology of utility assessment. IEEE Transactions on Systems, Man, and Cybernetics, 1977, SMC-7, 311-325.

Kazanowski, A.D. A standardized approach to cost effectiveness evaluation. In: J.M. English (Ed.) Cost-Effectiveness. New York: John Wiley, 1968.

Law, A.M., & Kelton, W.D. Simulation modeling and analysis. New York: McGraw-Hill, 1982.

Lev, B. Accounting and information theory. Chicago: American Accounting Association, 1969.

Lord, F.M., & Novick, M.R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Shannon, C.E. A mathematical theory of communication. Bell System Technical Journal, 1948, 27, pp. 379-423; 623-656.

Shannon, C.E., & Weaver, W. The mathematical theory of communication. Urbana, Ill.: University of Illinois Press, 1963.

Shannon, R.E. Systems Simulation: The Art and Science. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Shelnutt, J.B., Smillie, R.J., & Bercos, J. A consideration of Army training device proficiency assessment capabilities (ARI Technical Report TR-78-A20). Alexandria, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, June 1978.

Shubik, M., & Brewer, G.D. Models, simulations, and games: A survey. Santa Monica, Ca.: The Rand Corporation, May 1972.

Spartanics, Ltd. WEAPONEER: M16A1 Rifle Simulator. Rolling Meadows, Ill.: Author, 1976.

Stanley, J.C. Reliability. In R.L. Thorndike (Ed.), Educational measurement. Washington, D.C.: American Council on Education, 1971.

Thiel, H. Economics and information theory. Chicago: Rand McNally and North Holland Publishing Co., 1967.

Torgerson, W.S. Theory and methods of scaling. New York: John Wiley, 1958.

U.S. Army Combat Developments Command. ASARS Battle Model: Executive Summary. Ft. Monroe, Va.: Author, 1973.

U.S. Army Combat Developments Command. ASARS Battle Model: Narrative Description. Ft. Monroe, Va.: Author, 1973.

U.S. Army Training and Doctrine Command. Cost and operational effectiveness handbook (TRADOC PAM 11-8). Ft. Monroe, Va.: Author, November 1972.

U.S. Army Training and Doctrine Command. Cost and training effectiveness analysis handbook (Draft) (TRADOC PAM 71-10). Ft. Monroe, Va.: Author, July 1972.

Weitzman, D.O., Fineberg, M.L., Gade, P.A., & Compton, G.L. Proficiency maintenance and assessment in an instrument flight simulator. Human Factors, 1979, 21(6), 701-710.

Wood, R.O. Personal communication, 15 June 1981.

APPENDIX A

User Materials for In-House Rating Methods Study

INTRODUCTION

In the previous study (DORAC I) a methodological framework was developed for the examination of DORAC alternatives. The model developed for the study employed a set of mathematical rules for the assignment of quasi-quantitative values to the subjective judgment of experts. The technique used was a successive comparisons technique labeled MAUM (Multi-Attribute Utility Measurement).

The MAUM technique appeared to provide adequate information about DORAC alternatives, but it was time concuming for decision makers to implement, and the validity of the model could not be totally proven. It was, therefore, decided that other sets of decision rules would be examined and tested in conjunction with the MAUM. These rules could be used to check the validity of the model and for constructing alternative techniques which are easier to apply, or provide more objective information.

Four approaches have been developed for the creation of a judgmental utility scale. They are: 1) ranking, 2) partial paired comparisons, 3) rating, and 4) successive comparisons (the current MAUM).

The utility of a task or subtask is only a part of the "worth" of the task. Utility is defined as the "usefulness" of the task for providing information about the system. Utility does not tell how dependable the task can be measured, or how frequent the task is measured, its primary concern is the efficacy with which the task is applied to the particular system being evaluated.

Ranking

The ranking method is the simplest and most direct of all the methods to apply. The major assumption for this method is that the underlying distribution is essentially rectangular and the actual values are somewhat equidistant in the interval (implying that there are no extreme values). The ranking method normally makes a better showing on tests of internal consistency than that of paired comparisons, rating or successive comparisons, and is usually more time efficient. Results have shown that ranked data can be extremely valid when the scale values are correlated with objective criteria.

Procedures

Rank the tasks (subtasks and skills) within each level of the hierarchy. Do not cross over between task, subtasks and skills, and do not rank across sublevels of different tasks. The most useful task should be a 1. The next most useful task should be a 2 and so on...

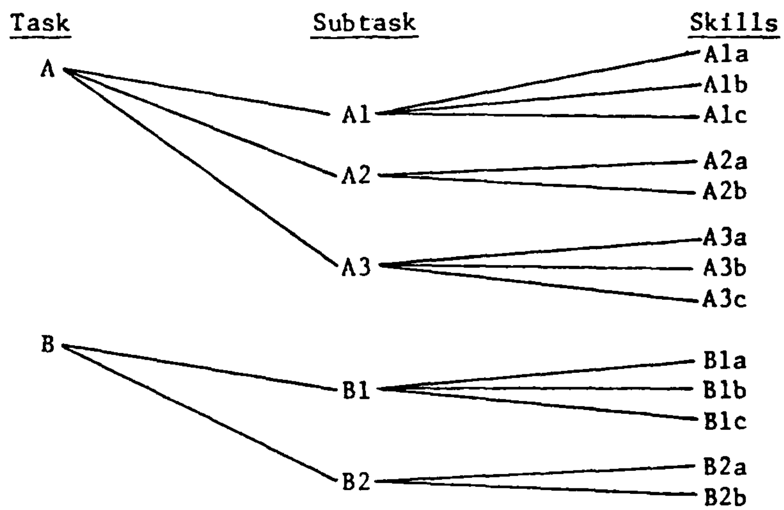


Figure 1

1. Rank order the tasks A, B, C.
2. Separately rank the following groups of subtasks:
    Subtasks A1, A2, A3.
    Subtasks B1, B2.
3. Separately rank the following groups of skills:
    Skills A1a, A1b, A1c.
    Skills A2a, A2b.
    Skills A3a, A3b, A3c.
    Skills B1a, B1b, B1c.
    Skills B2a, B2b.

Example:

Rank order within each block as shown.

Tasks	Rank	Subtask	Rank	Skills	Rank
A	2				
		A1	2		
				A1a	1
				A1b	2
				A1c	3
		A2	1		
				A2a	1
				A2b	2
		A3	3		
				A3a	3
				A3b	2
				A3c	1
B	1				
		B1	2		
				B1a	2
				B1b	1
				B1c	3
		B2	1		
				B2a	1
				B2b	2

Figure 2

Quantifying the values:

For each group of ranks:

1. Invert the ranks (take the highest numbered rank - biggest number - and add one to it, then subtract each individual rank from this value).
2. Normalize the rank values (total the ranks, then divide the individual ranks by the total).
3. Multiply the values across (or down) the hierarchy of tasks, subtasks, and skills until there is a single value for each of the skills. See Figure 3.

Task	Rank	Inverted Rank	Normalized Value	Subtask	Rank	Inverted Rank	Normalized Value	Skill	Rank	Inverted Rank	Normalized Value	Final Utility
A	2	1	.333									
				A1	2	2	.333					
								A1a	1	3	.500	.055
								A1b	2	2	.333	.037
								A1c	3	1	.167	.019
				A2	1	3	.500					
								A2a	1	2	.667	.111
								A2b	2	1	.333	.055
				A3	3	1	.167					
								A3a	3	1	.167	.009
								A3b	2	2	.333	.019
								A3c	1	3	.500	.028
B	1	2	.667									
				B1	2	1	.333					
								B1a	2	2	.333	.074
								B1b	1	3	.500	.111
								B1c	3	1	.167	.037
				B2	1	2	.667					
								B2a	1	2	.667	.297
								B2b	2	1	.333	.148

Figure 3

Rating

The rating method has several advantages over the ranking and paired comparison methods: 1) ratings require less time than either the paired comparisons or ranking methods, 2) ratings appear to be simpler for the "naive" individual who has a minimum of training, 3) ratings can be used with a larger number of items, 4) some investigators believe that the best judgments are given when each item is presented singularly (as is done with a rating scale) rather than pair wise or in a group (ranking), and 5) from investigation, it has been found that the judges perception of reliability is higher for rating than paired comparison or ranking.

Procedures

The method of rating consists of presenting a continuous scale marked off in units 0 - 100 to a judge. The judge is then asked to indicate on the scale his perceived value of the "usefulness" of the task (subtask or skill). The judge may select points between the groups of tens and may assign more than one criterion to a single position.

Quantifying the values:

1. Divide all rated values by 100.
2. Multiply the values across (or down) the hierarchy of tasks, subtasks and skills until there is a single value for each skill.
3. When all the skill values are determined, normalize the set of all skill values by totaling the set of values and dividing each value by the total. See Figure 4.

Task	Rating	Rating/100	Subtask	Rating	Rating/100	Skill	Rating	Rating/100	Uncorrected Utility Score	Normalized Utility Score
A	35	.35								
			A1	25	.25					
						A1a	40	.50	.044	.035
						A1b	25	.25	.022	.017
						A1c	15	.15	.013	.010
			A2	70	.70					
						A2a	70	.70	.172	.136
						A2b	50	.50	.123	.097
			A3	10	.10					
						A3a	40	.40	.014	.011
						A3b	50	.50	.018	.014
						A3c	75	.75	.026	.021
B	80	.80								
			B1	40	.40					
						B1a	50	.50	.160	.127
						B1b	90	.90	.288	.228
						B1c	30	.30	.096	.076
			B2	60	.60					
						B2a	40	.40	.192	.152
						B2b	20	.20	.096	.076
							Total		1,264	1.000

Figure 4

Partial Paired Comparison

Partial Paired Comparisons give results very similar to those of the ranking methods if the judges are consistent between the pairs (i.e., if task A importance is greater than task B, and task B importance is greater than task C importance). The partial paired comparison method is also similar to the ranking method in that the results can be extremely valid when correlated with objective criteria. The drawback of this method is that it can be very time consuming and wearying to the judges if a large number of pairs are being evaluated.

In the development of the Partial Paired Comparisons the measures are structured into a partial matrix. The judges are then asked to indicate between each set of two measures which is the more important (if they are equal then each measure is counted as one-half).

Procedures

1. Set up a partial Paired matrix for the tasks, subtasks and skills. To set up a partial matrix
    1) decide what measures are grouped together.
    2) create a row of the measures along the left side and a column of the measures along the top.
    3) then create a set of blocks that is a combination of all rows and columns.

    Example:

	Task 1	Task 2	Task 3	Task 4
Task 1	-			
Task 2	X	-		
Task 3	X	X	-	
Task 4	X	X	X	-

    4) Along the diagonal put a set of dashes.
    5) Use only the blocks above the dashes. (X out the other blocks)
2. Indicate in each block of the matrix the more important measure of the pair (if they are equal then put both numbers in the block and count each measure as one-half).

Quantifying the values:

1. For each individual matrix
   a) Count the number of times each task (or subtask or skill) occurs. If there is a tie between two measures count each one-half.
   b) Add one to each task count.
   c) Normalize the task count by totaling all the counts in the matrix and dividing the individual counts by the total
2. When all the matrices are evaluated according to the above procedures, multiply the normalized values across (down) the hierarchy of tasks, subtasks and skills until there is a single value for each skill.

MATRIX OF TASKS

	A	B
A	-	B
B		-

$\Sigma A = 0$

$\Sigma B = 1$

MATRICES OF SUBTASKS

	A1	A2	A3
A1	-	A2	A1
A2		-	A2
A3			-

$\Sigma A1 = 1$

$\Sigma A2 = 2$

$\Sigma A3 = 0$

	B1	B2
B1	-	B2
B2		-

$\Sigma B1 = 0$

$\Sigma B2 = 1$

MATRICES OF SKILLS

	A1a	A1b	A1c
A1a	-	A1a	A1a
A1b		-	A1b
A1c			-

$\Sigma A1a = 2$

$\Sigma A1b = 1$

$\Sigma A1c = 0$

	B1a	B1b	B1c
B1a	-	B1b	B1a
B1b		-	B1b
B1c			-

$\Sigma B1a = 1$

$\Sigma B1b = 2$

$\Sigma B1c = 0$

	A2a	A2b
A2a	-	A2a
A2b		-

$\Sigma A2a = 1$

$\Sigma A2b = 0$

	B2a	B2b
B2a	-	B2a
B2b		-

$\Sigma B2a = 1$

$\Sigma B2b = 0$

	A3a	A3b	A3c
A3a	-	A3b	A3c
A3b		-	A3c
A3c			-

$\Sigma A3a = 0$

$\Sigma A3b = 1$

$\Sigma A3c = 2$

Figure 5

Task	Task Count	Task Count +1	Normalized Task	Subtask	Subtask Count	Subtask Count +1	Normalized Subtask	Skill	Skill Count	Skill Count +1	Normalized Skill	Utility Value
A	0	1	.333									
				A1	1	2	.333					
								A1a	2	3	.500	.055
								A1b	1	2	.333	.037
								A1c	0	1	.167	.019
				A2	2	3	.500					
								A2a	1	2	.667	.111
								A2b	0	1	.333	.055
				A3	0	1	.167					
								A3a	0	1	.167	.009
								A3b	1	2	.333	.019
								A3c	2	3	.500	.028
B	1	2	.667									
				B1	0	1	.333					
								B1a	1	2	.333	.074
								B1b	2	3	.500	.111
								B1c	0	1	.167	.037
				B2	1	2	.667					
								B2a	1	2	.667	.297
								B2b	0	1	.333	.148

Figure 6

Tactical Employment of STINGER Weapon

On page 2 and 3 is a task structure for the Tactical Employment of the STINGER weapon for your reference. Page 4 and 5 are instructions for comparing tasks. You are requested to complete the tables of comparisons under the following Utility Dimensions.

Training Utility Dimension

1. Ranking procedures
2. Rating procedures
3. Partial Paired Comparisons procedure
4. Successive Comparisons procedure

Readiness Evaluation

1. Ranking procedures
2. Rating procedures
3. Partial Paired Comparisons procedures
4. Successive Comparisons procedures

Rank the following groups of tasks, with 1 being the highest.

Level 1

_____ (A) Preparation for combat

_____ (B) Defend against hostile aerial targets

Level 2

Group A Rank these five tasks

_____ (A1) Remove gripstock from launch tube

_____ (A2) Attach gripstock to launch tube

_____ (A3) Prepare basic load for tactical transport

_____ (A4) Load weapons in containers on M-416 trailer

_____ (A5) Secure missile container to M-416 trailer

Group B Rank these two tasks

_____ (B1) Prepare firing positions

_____ (B2) Target engagement procedures

Level 3

Group B1 Rank these four tasks

_____ (B1a) Select primary and alternative firing positions

_____ (B1b) Camouflage/conceal defensive positions

_____ (B1c) Erect camouflage screen and screen support system

_____ (B1d) Camouflage/conceal self and individual equipment

Group B2 Rank these six tasks

_____ (B2a) Operate TADDS

_____ (B2b) Perform observer procedures

_____ (B2c) Visually recognize aircraft

_____ (B2d) Exercise fire control of STINGER team

_____ (B2e) Use visual and hand signals to control STINGER team

_____ (B2f) Engage targets with STINGER

Level 4

Group B2F Rank these five tasks

_____ (B2f1) Direct defense of mobile/stationary assets from a march column

_____ (B2f2) Defend mobile/stationary assets from a march column

_____ (B2f3) Direct defense of a stationary asset from a prepared position

_____ (B2f4) Defend stationary assets from a prepared position

_____ (B2f5) Perform emergency procedures for hangfires, misfires and dud missiles

Rate on a scale 0-100 the following groups of tasks.

Level 1

_____ (A) Preparation for combat

_____ (B) Defend against hostile aerial targets

Level 2

Group A Rate these five tasks

_____ (A1) Remove gripstock from launch tube

_____ (A2) Attach gripstock to launch tube

_____ (A3) Prepare basic load for tactical transport

_____ (A4) Load weapons in containers on M-416 trailer

_____ (A5) Secure missile container to M-416 trailer

Group B Rate these two tasks

_____ (B1) Prepare firing positions

_____ (B2) Target engagement procedures

Level 3

Group B1 Rate these four tasks

_____ (B1a) Select primary and alternative firing positions

_____ (B1b) Camouflage/conceal defensive positions

_____ (B1c) Erect camouflage screen and screen support system

_____ (B1d) Camouflage/conceal self and individual equipment

Group B2 Rate these six tasks

_____ (B2a) Operate TADDS

_____ (B2b) Perform observer procedures

_____ (B2c) Visually recognize aircraft

_____ (B2d) Exercise fire control of STINGER team

_____ (B2e) Use visual and hand signals to control STINGER team

_____ (B2f) Engage targets with STINGER

Level 4

Group B2f Rate these five tasks

_____ (B2f1) Direct defense of mobile/stationary assets from a march column

_____ (B2f2) Defend mobile/stationary assets from a march column

_____ (B2f3) Direct defense of a stationary asset from a prepared position

_____ (B2f4) Defend stationary assets from a prepared position

_____ (B2f5) Perform emergency procedures for hangfires, misfires and dud missiles

Compare the task in the column to the task in the row for each open block. Within the block place the task ID which you consider to be the most useful of the two. If they are the same put both task IDs in the block.

Level 1

Task I.D.	Preparation for Combat (A)	Defend Against Hostile Aerial Targets (B)
Preparation for Combat (A)	X	
Defend Against Hostile Aerial Targets (B)	X	X

Level 2

Group A

Task	Task I.D.	Remove Gripstock from Launch Tube	Attach Gripstock to Launch Tube	Prepare Basic Load for Tactical Transport	Load Weapons in Containers on M-416 Trailer	Secure Missile Container to M-416 Trailer
Remove Gripstock from Launch Tube	(A1)	X				
Attach Gripstock to Launch Tube	(A2)	X	X			
Prepare basic load for Tactical Transport	(A3)	X	X	X		
Load Weapons in Containers on M-416 Trailer	(A4)	X	X	X	X	
Secure Missile Container to M-416 Trailer	(A5)	X	X	X	X	X

Group B

Task	Task I.D.	Prepare Firing Positions (B1)	Target Engagement Procedures (B2)
Prepare Firing Positions	(B1)	X	
Target Engagement Procedures	(B2)	X	X

Level 3

Group B1

Task	Task I.D.	Select Primary and Alternative Firing Positions (B1a)	Camouflage/Conceal Defensive Positions (B1b)	Erect Camouflage Screen and Screen Support System (B1c)	Camouflage/Conceal Self and Individual Equipment (B1d)
Select Primary and Alternative Firing Positions	(B1a)	X			
Camouflage/Conceal Defensive Positions	(B1b)	X	X		
Erect Camouflage Screen and Screen Support System	(B1c)	X	X	X	
Camouflage/Conceal Self and Individual Equipment	(B1d)	X	X	X	X

Level 3

Group B2

	Task I.D.	Operate TADDS (B2a)	Perform Observer Procedures (B2b)	Visually Recognize Aircraft (B2c)	Exercise Fire Control of STINGER Team (B2d)	Use Visual & Hand Signals to Control STINGER Team (B2e)	Engage Targets with STINGER (B2f)
Operate TADDS	(B2a)	X					
Perform Observer Procedures	(B2b)	X	X				
Visually Recognize Aircraft	(B2c)	X	X	X			
Exercise Fire Control of STINGER Team	(B2d)	X	X	X	X		
Use Visual & Hand Signals to Control STINGER Team	(B2e)	X	X	X	X	X	
Engage Targets with STINGER	(B2f)	X	X	X	X	X	X

Level 4

Group B2f

Task I.D.	Direct Defense of Mobile/ Stationary Assets from a March Column (B2f1)	Defend Mobile/Stationary Assets from a March Column (B2f2)	Direct Defense of a Stationary Asset from a Prepared Position (B2f3)	Defend Stationary Assets from a Prepared Position (B2f4)	Perform Emergency Procedures for Hangfires, Misfires and Dud Missiles (B2f5)
Direct Defense of Mobile/ Stationary Assets from a March Column (B2f1)	X				
Defend Mobile/Stationary Assets from a March Column (B2f2)	X	X			
Direct Defense of a Stationary Asset from a Prepared Position (B2f3)	X	X	X		
Defend Stationary Assets from a Prepared Position (B2f4)	X	X	X	X	
Perform Emergency Procedures for Hangfires, Misfires and Dud Missiles (B2f5)	X	X	X	X	X

Task Level 1	Level 2	Level 3	Level 4
(A) Preparation for Combat	(A1) Remove Gripstock from Launch Tube		
	(A2) Attach Gripstock to Launch Tube		
	(A3) Prepare Basic Load for Tactical Transport		
	(A4) Load Weapons in Containers on M-416 Trailer		
	(A5) Secure Missile Containers to M-416 Trailer		
(B) Defend Against Hostile Aerial Targets	(B1) Prepare Firing Positions	(B1a) Select Primary and Alternative Firing Positions	
		(B1b) Camouflage/Conceal Defensive Positions	
		(B1c) Erect Camouflage Screen and Screen Support System	
		(B1d) Camouflage/Conceal Self and Individual Equipment	

Task Level 1	Level 2	Level 3	Level 4
(B) Defend Against Hostile Aerial Targets (Cont'd)	(B2) Target Engagement Procedures	(B2a) Operate TADDS	
		(B2b) Perform Observer Procedures	
		(B2c) Visually Recognize Aircraft	
		(B2d) Exercise Fire Control of STINGER Team	
		(B2e) Use Visual & Hand Signals to Control Fire of STINGER Team	
		(B2f) Engage Targets with STINGER	(B2f1) Direct Defense of Mobile/Stationary Assets from a March Column
			(B2f2) Defend Mobile/Stationary Assets from a March Column
			(B2f3) Direct Defense of Stationary Assets from a Prepared Position
			(B2f4) Defend Stationary Assets from a Prepared Position
			(B2f5) Perform Emergency Procedures for Hangfires, Misfires, and Dud Missiles

Level 1 Group ____	Initial Rating (from Step 2)	Revised Ratings (Iterations of Step 3)				Importance Weight (From Step 4)
		1	2	3	4	
(A) Preparation for Combat						
(B) Defend Against Hostile Aerial Targets						

Level 2 Group A	Initial Rating (from Step 2)	Revised Ratings (Iterations of Step 3)				Importance Weight (From Step 4)
		1	2	3	4	
A1 Remove Gripstock from Launch Tube						
A2 Attach Gripstock to Launch Tube						
A3 Prepare Basic Load for Tactical Transport						
A4 Load weapons in Containers on M-416 Trailer						
A5 Secure Missile Containers to M-416 Trailer						

Level 2 Group B	Initial Rating (from Step 2)	Revised Ratings (Iterations of Step 3)				Importance Weight (from Step 4)
		1	2	3	4	
B1 Prepare Firing Positions						
B2 Target Engagement Procedures						

Level 3 Group B1	Initial Rating (from Step 2)	Revised Ratings (Iterations of Step 3)				Importance Weight (from Step 4)
		1	2	3	4	
B1a Select Primary and Alternative Firing Positions						
B1b Camouflage/conceal Defensive Positions						
B1c Erect Camouflage Screen and Screen Support System						
B1d Camouflage/conceal Self and Individual Equipment						

Level 3 Group B2	Initial Rating (from Step 2)	Revised Ratings (Iterations of Step 3)				Importance Weight (from Step 4)
		1	2	3	4	
B2a Operate TADDS						
B2b Perform Observer Procedures						
B2c Visually Recognize Aircraft						
B2d Exercise Fire Control of STINGER Team						
B2e Use Visual and Hand Signals to Control Fire of STINGER Team						
B2f Engage Targets with STINGER						

Level 4 Group B2f	Initial Rating (from Step 2)	Revised Ratings (Iterations of Step 3)				Importance Weight (from Step 4)
		1	2	3	4	
B2f1 Direct Defense of Mobile/Stationary Assets from a March Column						
B2f2 Defend Mobile/ Stationary Assets From a March Column						
B2f3 Direct Defense of Stationary Assets From a Prepared Position						
B2f4 Defend Stationary Assets From a Prepared Position						
B2f5 Perform Emergency Procedures for Hangfires, Misfires and Dud Missiles						

APPENDIX B

User Materials for Ft. Benning Formative Tryouts

Introduction

Applied Science Associates, Inc. (ASA), in conjunction with the Army Research Institute (ARI) Field Unit at Ft. Benning, has been working on a project concerned with the development of methods for using training devices in performance assessment. In the language of this project, a training device, or set of devices, used for performance assessment is referred to as a Training Device Operational Readiness Assessment Capability, or DORAC.

One of the issues in the DORAC project, and the one that we are asking you to help us address, concerns deciding what measurement capability to include with training devices used for performance assessment. Modifying a training device to facilitate performance measurement can be a costly undertaking. Care must be taken therefore to include in a DORAC only the capability to measure those performances that are truly useful to trainers and commanders. An analysis directed at identifying the most cost-effective performance measurement capability is termed, again in the language of the project, Cost and Information Effectiveness Analysis, or CIEA.

ASA and ARI have developed a set of procedures for the conduct of CIEA. We need to evaluate certain of these procedures in terms of their usability and the nature of the results they produce. The portions of the analysis that you are being asked to participate in concern establishing the worth of the performance status information obtainable from a DORAC. Information worth will be considered from two perspectives: Unit Readiness Evaluation and Unit Training Management. Information concerning a performance is defined to have worth for Unit Readiness Evaluation in direct proportion to the judged contribution of that performance to individual and/or unit combat effectiveness. Information regarding a performance is defined to have worth for Unit Training Management in direct relation to the extent to which you, as a trainer or commander, could and would make use of it to formulate or revise your individual and collective training plans.

The task we are asking you to do consists of five steps. In the first step, you will be asked to determine the comparative worth of each of the uses for DORAC information: Unit Readiness Evaluation and Unit Training Management.

Following this step, you will be asked to rate a series of M16A1 performances on their worth for Unit Readiness Evaluation. The third and fourth steps will focus upon the capabilities of the alternatives being considered in the analysis. Here, you will be asked to provide what we call Measurement Precision and System Effectiveness ratings. Finally, you will be asked to repeat Step 2, focusing this time upon the worth of performance status information for Unit Training Management. Each of the steps of the analysis will be described in greater detail before you are asked to carry it out. Ratings will be developed by group concensus with a single set of ratings for each step. Also, an ASA staff member will guide you through each step of the analysis. A tentative time schedule for the day's activities is provided on the next page.

Tentative Schedule

8:00 - 8:15	Introduction
8:15 - 8:30	Establish weights for Worth Dimensions
8:30 - 10:00	Rate utility of performances for Unit Readiness Evaluation
10:00 - 10:15	Break
10:15 - 12:00	Obtain Measurement Precision Ratings
12:00 - 1:30	Lunch
1:30 - 3:00	Obtain System Effectiveness ratings
3:00 - 3:15	Break
3:15 - 4:30	Rate utility of performances for Unit Training Management

STEP 1

Importance Weights for Worth Dimensions

Part 1:

Importance weights for Worth Dimensions (WDs) are assinged using the series of steps presented below. To assist in the rating process, a rating development sheet is provided on the next page.

1. Rank the WDs in order of importance in column B.
2. Rate the WDs on importance: (Column C)
   a. Assign the least important WD a rating of 10 in column C.
   b. Consider the next-least-important WD. How much more important is it than the least important? Assign it a number that reflects that ratio. For example, if the second-least-important WD is judged to be four times as important as the first, it is assigned a score of 40. Continue up through the list of WDs. Check each set of ratios as each new judgment is made.
   c. Review your ratings to insure that they reflect the actual importance of each of the WDs. Are the ratios of distances between WDs correct? Make any necessary adjustments in your ratings and list the results in column F.

Part 2:

A. If only two WDs are noted, sum the resulting scores. Divide each score by the resulting sum. Round to two places. Record results in column F, which completes this step.

B. If more than two (2) WDs are being rated, carry out the following additional series of steps to improve the reliability of the resulting importance weights, using column E, 1 to 10, for each repetition.

1. Compare the first (most important) WD with the remaining ones put together. Is it more important, equally important, or less important than all the others put together?
2. If the first WD is more important than all of the others put together, see if it's importance rating is greater than the sum of the importance ratings for all of the other WDs. If not, change the importance rating of the first WD so that it is greater than the sum of the others.

3. If the first WD is of equal importance to all the others put together, see if its importance rating is equal to the sum of the importance ratings of all the other WDs. If it is not, change the importance rating of the first WD so that it is equal to the sum of the others.
4. If the first WD is less important than all the others put together, see if its importance rating is less than the sum of the importance ratings of all of the other WDs. If it is not, change the importance rating of the first WD so that it is less than the sum of the others.
5. If the first WD was considered more important or equally important than all the others put together, apply the above procedure to the second-most-important WD on the list. Is it more important, equally important, or less important than all the other farther down the list put together? Then, proceed as in (2), (3), and (4) above, applying the revision procedure to the second WD instead of the first.
6. If the first WD was considered less important than all the others put together, compare the first WD with all the remaining ones put together, except the lowest rated one. Is the first WD more important, equally important, or less important than all of the others farther down the list except the lowest one put together? Then proceed as in (2), (3), and (4) above. If (2) or (3) are applicable, proceed to (5) after applying (2) or (3). If (4) is applicable, proceed as in this paragraph (6) again, comparing the first WD with all the remaining ones put together except the lowest two. As long as (4) is applicable, the procedures of this paragraph (6) are repeated until the first WD is compared with the second and third WDs put together. Then, even if (4) is still applicable, proceed to (5).
7. Continue the above procedure until the third-from-the-lowest WD has been compared with the two lowest WDs on the list.
8. Sum the resulting scores. Divide each score by the resulting sum. Round to two places. Record results in column F, which completes this step.

Date ______________

Group # ______________

Job ______________

STEP 1

Rating Development Sheet

A	B	C	D	E										F
				Revised Rating (Part 2)										
Elements	Rank	Initial Rating	Final Rating	1	2	3	4	5	6	7	8	9	10	Importance Weight
WD #1 Unit Readiness Evaluation														
WD #2 Unit Training Management														

Step 1,2,4B

STEP 2A

Utility Ratings for Performances

Utility scores for performances are obtained using the following series of steps. Use the attached sheet to record your results.

1. List the performances in descending order of utility in column A (i.e., value for decision-making) with respect to the WD being considered and rank them in column B.
2. If there are ten or fewer performances nested under a WD, obtain utility scores using the following substeps:
   a. Assign the least important performance a rating of 10 in column C.
   b. Consider the next-least-important performance. How much more important is it than the least important? Assign it a number that reflects that ratio. For example, if the second-least-important performance is judged to be four times as important as the first, it is assigned a score of 40. Continue up through the list of performances entering ratings in column C. Check each set of ratios as each new judgment is made.
   c. Review your ratings to insure that they reflect the actual utility of each of the performances. Are the ratios of distances between performances correct? Make any necessary adjustments to your ratings and record in column D.
3. If a WD includes more than ten performances, obtain utility scores as follows:
   a. Select one performance at random.
   b. Randomly assign each of the remaining performances to groups of approximately equal size, with no more than five performances to a group and record performances in column A of a separate sheet for each group.
   c. Add the performance selected in Substep (a) to each group and assign it a rating of 100 in column C. This index performance will serve to re-link each of the groups later (Substep e).

d. Rank each of the performances in each group in order of descending utility in column B. Then, assign numerical ratings to them following the procedure outlined in Step 2. Keep the rating of the performance selected in Substep (a) fixed at 100.

e. Transfer the initial ratings (column C) of all the performances to column C of the initial list. Compare these ratings with the initial rankings from Step 1. Note any differences in rankings. If the initial list is judged correct, repeat Substep (d) to adjust the affected groups and reconcile the evaluations, in column D.

Date ______________

Group # ______________

Job ______________

STEP 2A

Rating Development Sheet

A	B	C	D	E										F
				Revised Rating (Part 2)										
Elements	Rank	Initial Rating	Final Rating	1	2	3	4	5	6	7	8	9	10	Importance Weight
I. Prepare to Engage Targets														
1. Condition weapon to operate														
a. Calibrate weapon														
b. Load weapon														
2. Maintain weapon operation														
c. Reload weapon														
d. Reduce stoppage														
3. Operate weapon														
e. Soldier-weapon interface														
f. Marksmanship factors														
II. Engage Targets Individually														
4. Assault mode														
g. Aimed fire														
h. Pointed fire														
5. Defensive positions														
g. Aimed fire														
h. Pointed fire														

Step 1,2,4B

Date ______________

Group # ______________

Job ______________

STEP 2A

Rating Development Sheet

A	B	C	D	E										F
Elements	Rank	Initial Rating	Final Rating	Revised Rating (Part 2)										Importance Weight
				1	2	3	4	5	6	7	8	9	10	
II. Engage Targets (Cont'd)														
6. Patrol operations														
g. Aimed fire														
h. Pointed fire														
7. Detect and identify targets														
III. Team Fire														
8. Conduct fire and maneuver														
i. Control fire team movement														
j. Move as a member of a fire team														
k. Move under direct fire														
9. Practice ammunition rationing														
l. Control rate of fire														
m. Control distribution of fire														

Step 1,2,4B

STEP 2B

Utility Ratings for Performances

Utility ratings for performances are obtained using the following series of steps. In developing your utility ratings, please use the attached sheet. Complete the steps separately for each level of the hierarchy.

1. Rank the performance statements within each of the levels of the hierarchy in order of their utility (i.e., value for decision-making) with respect to the WD being considered and record rank in column B. Do not corss over levels of the hierarchy in assigning your ranks.
2. If there are ten (10) or fewer performances at a given level:
   a. Assign the least important performance a rating of 10 in column C.
   b. Consider the next-least important performance. How much more important is it than the least important? Assign it a number that reflects that ratio. For example, if the second-least-important performance is judged to be four times as important at the first, it is assigned a score of 40. Continue up through the list of performances entering rates in column C. Check each set of ratios as each new judgment is made.
   c. Review your ratings to insure that they reflect the actual utility of each of the performances. Are the ratios of distances between performances correct? Make any necessary adjustments to your ratings and record in column D.
3. If there are more than ten (10) performances at a given level:
   a. Select one performance at random.
   b. Randomly assign each of the remaining performances to groups of approximately equal size, with five to seven performances to a group and, on a separate rating sheet for each group, record performances in column D.

c. Add the performance selected in Substep (a) to each group and assign it a rating of 100 in column C. This index performance will serve to re-link each of the groups later.

d. Rank each of the performances in each group in terms of descending utility in column B. Then, assign numerical ratings to them following the procedure for fewer than ten performances, in column C. Keep the rating of the performance selected in Substep (a) fixed at 100.

e. Transfer initial ratings (column C) of all of the performances to column C of the initial list. Compare this list with the initial rankings. Note any differences in these ratings. If the initial list is judged correct repeat Substep (e) to adjust the affected groups and reconcile the evaluations in column D.

Date ______________

Group # ______________

Job ______________

STEP 2B Level 1

Rating Development Sheet

A	B	C	D	E										F
				Revised Rating (Part 2)										
Elements	Rank	Initial Rating	Final Rating	1	2	3	4	5	6	7	8	9	10	Importance Weight
I. Prepare to Engage Targets														
II. Engage Targets Individually														
III. Team Fire														

Step 1,2,4B

Date ____________

Group # ____________

Job ____________

STEP 2B Level 2

Rating Development Sheet

A	B	C	D	E										F
Elements	Rank	Initial Rating	Final Rating	Revised Rating (Part 2)										Importance Weight
				1	2	3	4	5	6	7	8	9	10	
I. Prepare to Engage Targets														
1. Condition weapon to operate														
2. Maintain weapon operation														
3. Operate weapon														
II. Engage Targets Individually														
4. Assault mode														
5. Defensive position														
6. Patrol operation														
7. Detect and identify targets														
III. Team Fire														
8. Conduct fire and maneuver														
9. Practice ammunition rationing														

Step 1,2,4B

Date ____________

Group # ____________

Job ____________

STEP 2B Level 3

Rating Development Sheet

A	B	C	D	E										F
Elements	Rank	Initial Rating	Final Rating	Revised Rating (Part 2)										Importance Weight
				1	2	3	4	5	6	7	8	9	10	
I. Prepare to Engage Targets														
1. Condition weapon to operate														
a. Calibrate weapon														
b. Load weapon														
2. Maintain weapon operation														
c. Reload weapon														
d. Reduce stoppage														
3. Operate weapon														
e. Soldier-weapon interface														
f. Marksmanship factors														

Step 1,2,4B

Date ____________

Group # ____________

Job ____________

STEP 2B Level 3

Rating Development Sheet

A	B	C	D	E										F
Elements	Rank	Initial Rating	Final Rating	Revised Rating (Part 2)										Importance Weight
				1	2	3	4	5	6	7	8	9	10	
II. Engage Targets Individualls														
4. Assault mode														
g. Aimed fire														
h. Pointed fire														
5. Defensive mode														
g. Aimed fire														
h. Pointed fire														
6. Patrol operation														
g. Aimed fire														
h. Pointed fire														
III. Team Fire														
8. Conduct fire and maneuver														
i. Control fire team movement														
j. Move as a member of a fire team														
k. Move under direct fire														

Step 1,2,4B

Date ____________

Group # ____________

Job ____________

STEP 2B Level 3

Rating Development Sheet

A	B	C	D	E										F
Elements	Rank	Initial Rating	Final Rating	Revised Rating (Part 2)										Importance Weight
				1	2	3	4	5	6	7	8	9	10	
III. Team Fire (Cont'd)														
9. Practice ammunition rationing														
1. Control rate of fire														
m. Control distribution of fire														

Step 1,2,4B

Training Devices and (DORAC) Alternatives

The DORAC alternatives selected for use in the demonstration analysis are concerned with the assessment of marksmanship proficiency. The current method used to assess marksmanship proficiency is Record Fire (RF) conducted one time per year. During RF, each soldier is taken to a firing range and assessed in a 40-round live-fire exercise. Prone and foxhole firing positions are employed; range and target exposure time also vary. A candidate is rated as "qualified" if he/she achieves 17 (23 at Ft. Benning) or more hits out of the 40 possible. Figure 1 presents the firing positions, target ranges, and times used in RF.

The training device undergoing evaluation as an adjunct to RF is the Weaponeer. Weaponeer (WP) is an M16A1 remedial marksmanship trained designed to isolate individual performance deficiencies. A simulated M16A1 rifle is equipped with a target sensor and each target contains a light emitting diode (LED) which is sensed by the target sensor on the rifle. A predicted round impact point is determined by the LED-target sensor alinement. WP has a memory for recording up to 32 predicted shot impacts and a printer for providing a printout of all shots on selected targets. Rifle recoil is simulated with recoil energy being variable from no recoil to a recoil intensity 40 percent greater than the recoil of a standard M16A1 rifle. Three types of magazines are provided for use with the rifle: a 20-round (unlimited fire) magazine, a 30-round (unlimited fire) magazine, and a limited fire magazine that allows from 1 to 30 simulated rounds in the magazine. A headset is provided for simulating the firing sound of an M16A1 rifle. The WP also includes a selection for random misfire.

WP can present three targets: a scaled 25 meter zeroing target, a scaled 100 meter 'E' type silhouette target, and a 250 meter 'E' type silhouette target. Any target selected can be raised at random during a 1 to 9 second time frame and can remain in a raised position for a duration of 2, 3, 5, 10, 15, 20, 25 seconds, or continuous. The WP provides a target 'Kill' function: a selection that will cause a raised target to drop when it is hit. Firing pads used with the WP provide the capability for the user to fire from the foxhole or prone positions.

RECORD FIRE SCORECARD

Last Name, FI, MI	SSAN	Unit	Date	Time

TABLE 1. FOXHOLE POSITION

RD	Range (M)	Time (Sec)	Hit	Miss	No Fire
1	50	5			
2	200	5			
3	100	5			
4	150	5			
5	300	10			
6	250	10			
7	50	5			
8	200	5			
9	150	5			
10	250	10			
		TOTAL			

TABLE 3 PRONE POSITION

RD	Range (M)	Time (Sec)	Hit	Miss	No Fire
1	150	5			
2	300	10			
3	100	10			
4	200				
5	150	15			
6	250				
7	100	15			
8	300				
9	200	15			
10	300				
		TOTAL			

TABLE 2 PRONE POSITION

RD	Range (M)	Time (Sec)	Hit	Miss	No Fire
1	100	15			
2	250				
3	300	10			
4	50	10			
5	200				
6	150	15			
7	300				
8	50	10			
9	200				
10	100	5			
		TOTAL			

TABLE 4 FOXHOLE POSITION

RD	Range (M)	Time (Sec)	Hit	Miss	No Fire
1	100	10			
2	200				
3	250	20			
4	300				
5	100	15			
6	250				
7	250	10			
8	300	10			
9	50	10			
10	100				
		TOTAL			

TABLE	HIT	MISS
1		
2		
3		
4		
TOTAL		

SCORER S SIGNATURE & LANE NO

QUALIFICATION SCORES AND RATING

POSSIBLE	40
EXPERT	28 40
SHARPSHOOTER	24 27
MARKSMAN	17 23
UNQUALIFIED	16 BELOW

QUALIFICATION (CIRCLE ONE)

EXPERT, SHARPSHOOTER, MARKSMAN, UNQUALIFIED

FIRER S QUALIFICATION SCORE ______________

OFFICER S SIGNATURE

83910

A video display allows an observer to monitor individual shots and replay the last 3 seconds of each of the first 3 shots. Scoring available with the WP video display includes: the target on display, the number of hits on the target, the number of misses, late shots (fired after target drops), the shot number, and the total number of shots fired.

Methods and training devices used alone will not always constitute DORAC alternatives. In fact, DORAC alternatives will usually consist of sets of training devices/methods used in combination and a usage scenario. For the demonstration analysis, the following devices/methods and usage scenarios constitute the DORAC alternatives that you are to evaluate:

1. RF conducted one time per year [RF(1)].
2. RF twice per year (every six months) [RF(2)].
3. RF quarterly [RF(4)].
4. WP once per year [WP(1)].
5. RF once, WP once (every six months)[RF(1) + WP(1)].
6. RF once, WP three times [RF(1) + WP(3)].

Orientation for DORAC Data Collection
To Determine Information Worth
For Determining Unit Readiness

1. The purpose of the collection effort is to identify those tasks whose measurement provides effective, accurate indications of unit mission performance capability. Estimations of task importance or worth must be made on the basis of the value of the information obtained, rather than the inherent value of the task itself. For example, the task "Identify Enemy Vehicles" is without doubt an important and valuable task. For an armor crewman, it is a mission-essential task; however, its mission value to a member of a general support maintenance unit is somewhat questionable. Another example is "Prepare Forms and Requests". This task has little mission-relevance to an Infantryman; but, it is mission-essential to the S-4 clerk who prepares ammunition requisitions to supply the Infantryman. Tasks then must be considered both from the view of unit, or collective, task relationship, and from the context of job relationship to unit mission.

2. The attitudes of data collection participants will greatly affect the accuracy and validity of information developed.

a. Tasks must be evaluated from the view of "Should they be evaluated?", not "Can they be evaluated?". The concern here is the need for evaluation, not the capability to measure.

b. Task measurement worth is dependent upon its value to estimating unit mission performance, not unit level of training. While training efficiency may be a side product of DORAC, it is not the primary goal.

c. Task measurement worth must not consider individual proficiency as an end result; again, the need is to measure unit mission capability.

d. Whether or not a unit mission, or task is currently practiced in training must not be a consideration in determining task information worth. The concern is the need to measure, not the importance to current operating procedures.

e. The value of task measurement must be determined on the basis of task relationship to unit mission accomplishment, not unit appearance or garrison functions. Ability to operate a vehicle in a mounted review is just not as important as operating the vehicle across rough terrain. The unit mission tasks which need to be accessed are combat, not garrison, missions. This does not mean that administrative tasks are not important; it does mean that their importance is determined by their relationship to estimating unit combat mission performance.

f. The same task may appear under different missions, and must be evaluated in the context of each mission.

Orientation for DORAC Data Collection
To Determine Information Worth
For Training Management

1. In the earlier steps, you were asked to develop ratings for the worth of information in assessing Unit Readiness. During this step, you will be repeating step two, developing utility ratings for performance; except that this time you must weigh the utility value for use in Training Management. The primary concern here is: Will the information provided by measuring the performance affect training plans and decisions? View each performance from the following prospectives:

a. Whether or not a performance is currently being trained is not a consideration. The concern is the need to measure the performance regardless of current operating procedures.

b. Does the results of measuring the performance indicate level of proficiency and training needs?

c. Will a change in the results of performance measurement cause changes in training plans, decisions or procedures. Results that do not affect future training are of little value.

2. The utility of information for training management is primarily based on its relationship to decision making. If training programs, plans, and allocation of training resources may be changed by the results of performance measurement, then it should have a high utility rating for Training Management. If either success or failure of the performance measured will not alter the training situation, it's utility is low.

STEP 3A

Measurement Precision

Refer to the attached worksheet. For each block containing an entry (i.e., performance measurement is possible), rate the measurement method on the precision of the data it will provide in column A. Factors that should be considered in assigning measurement precision ratings include:

1. The judged reliability of the method. Reliability is defined as the extent to which a measurement method provides accurate and stable perforamnce scores.
2. The validity of the resulting performance scores. Validity is defined as the extent to which an evaluatee's score, as defined through the operational performance measure, is a true representation of the performance it is intended to measure.

Measurement Precision ratings are to be assigned according to the following scale:

0 - Zero Precision; no reliability, no validity.

25 - Low Precision; low reliability and low validity.

50 - Moderate Precision; acceptable validity, moderately high reliability.

75 - High Precision; high validity and high reliability.

100 - Perfect Precision; perfect validity and perfect reliability.

Date ______

Group # ______

Job ______

STEP 3A

Measurement Precision Worksheet

Performance Measure	Alternative 1 — Device: Record Fire; OPM Measurement Method	A	B	C	D	Alternative 2 — Device: Weaponeer; OPM Measurement Method	A	B	C	D
I. Prepare to Engage Targets										
1. Condition Weapon to Operate	Correct impact point; group									
A. Calibrate (zero) Weapon	Evaluator records group					0				
1. Adjust for minute of angle										
2. Adjust for gravity effects	Evaluator records group					0				
3. Obtain proper sight-weapon alinement	Evaluator records group					0				
B. Load Weapon	Go-NoGo									
4. Inspect Magazine for top round alinement	Evaluator Checklist					0				
5. Apply proper intensity flow to seat magazine	Evaluator Checklist					Go-NoGo Evaluator checklist				
6. Insure first round is seated in chamber	Evaluator Checklist					0				

Step 3,4A

Date ____________

Group # ____________

Job ____________

Measurement Precision Worksheet

Performance Measure	Alternative 1 — Device Record Fire — OPM Measurement Method	A	B	C	D	Alternative 2 — Device Weaponeer — OPM Measurement Method	A	B	C	D
2. Maintain Weapon Operation	Go-NoGo					Go-NoGo				
C. Reload weapon	Evaluator Observation					Evaluator Observation				
7. Remove magazine										
4. Same as above	Evaluator Observation					0				
5. Same as above	Evaluator Observation					Go-NoGo Evaluator Observ.				
6. Same as above	Evaluator Observation					0				
D. Reduce stoppage	Go-NoGo					Go-NoGo				
8. Perform immediate action procedure	Evaluator Checklist					Evaluator Checklist				
9. Inspect weapon and diagnose stoppage	Evaluator Checklist					0				
10. Take appropriate action for reducing stoppage	Evaluator Checklist					0				
3. Operate Weapon	Go-NoGo					Go-NoGo				
E. Soldier-weapon Interface	Evaluator Checklist					Evaluator Checklist Machine assisted				
11. Trigger control (squeeze)										
12. Grip on weapon	Evaluator Checklist					Evaluator Checklist				
13. Breath control	Evaluator Checklist					Evaluator Checklist				
14. Relaxation	Evaluator Checklist					Evaluator Checklist				

(Continued)

Step 3,4A

Date ______________

Group # ______________

Job ______________

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device Record Fire OPM Measurement Method	A	B	C	D	Alternative 2 Device Weaponeer OPM Measurement Method	A	B	C	D
15. Compensate for recoil	Evaluator Checklist					Evaluator Checklist				
16. Stock weld	Evaluator Checklist					Evaluator Checklist				
17. Firing elbow	Evaluator Checklist					Evaluator Checklist				
F. Marksmanship Factors 18. Compensate for range	Subjective Go-NoGo Observer records score for each target					Subjective Go-NoGo Evaluator verifies machine assisted				
19. Compensate for wind	Observer records score for each target					0				
20. Compensate for target movement	0					0				
21. Conduct transfer of fire	Subjective Go-NoGo Observer records score/behavior					Subjective Go-NoGo Observer records score/behavior				
22. Obtain proper sight picture	Go-NoGo M-16 sighting device					Go-NoGo M-16 sighting device, video playback				
23. Determine proper aim point	Subjective Go-NoGo Evaluator observes hit pattern					Go-NoGo Machine video playback				
24. Determine proper weapon aspect-round impact relationship	Subjective Go-NoGo Evaluator observes behavior					0				

Step 3,4A

Date ____________

Group # ____________

Job ____________

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device Record Fire OPM Measurement Method	A	B	C	D	Alternative 2 Device Weaponeer OPM Measurement Method	A	B	C	D
II. Engage Targets (In): 4. Assault Mode (using) g. Aimed fire (from) 25. Prone position	Minimum 17 hits/40 rounds (23 hits at Ft. Benning) Observer records hits (composite of a,c,&d)					Predetermined percentage of hits (variable) Machine recorded score printout, video playback				
26. Standing position	Not scored					"				
27. Foxhole supported position	See (a)					"				
28. Foxhole unsupported position	See (a)					"				
h. Pointed fire: 29. From standing position	Not currently assessed Observer records hits (composite of e & f)					0				
30. While moving	See (e)					0				
5. Defensive positions (using): g. Aimed fire (from) 25. Prone position	Same as 25 above)					(Same as 25 above)				
26. Standing position	0					"				
27. Foxhole supported position	See (a)					"				

Step 3,4A

Date ______________

Group # ______________

Job ______________

Measurement Precision Worksheet

Performance Measure	Alternative 1 — Device: Record Fire; OPM Measurement Method	A	B	C	D	Alternative 2 — Device: Weaponeer; OPM Measurement Method	A	B	C	D
28. Foxhole unsupported position	See (a)					(Same as 25 above)				
h. Pointed fire: 29. From standing position	(Same as 26 above)					0				
30. While moving	(See (e)					0				
6. Patrol Operation (using): g. Aimed fire (from): 25. Prone position	Minimum 17 hits/40 rounds (23 hits at Ft. Benning) Observer records hits (composite of a,c,&d)					Predetermined percentage of hits (variable) Machine recorded score Printout, video playback				
26. Standing position	Not scored					"				
27. Foxhole supported position	See (a)					"				
28. Foxhole unsupported position	See (a)					"				
h. Pointed fire: 28. From standing position	Not currently assessed Observer records hits (composite of e & f)					0				
29. While moving	See (e)					0				

Step 3,4A

Date ______

Group # ______

Job ______

Measurement Precision Worksheet

Performance Measure	Alternative 1: Device Record Fire; OPM Measurement Method	A	B	C	D	Alternative 2: Device Weaponeer; OPM Measurement Method	A	B	C	D
7. Detect and Identify Targets	0					0				
III. Team Fire.										
8. Conduct fire and maneuver	0					0				
i. Control fire team movement										
j. Move as a member of a fire team	0					0				
k. Move under direct fire	0					0				
9. Practice ammunition rationing										
l. Control rate of fire	0					0				
m. Control distribution of fire	0					0				

Step 3, 4A

Measurement Precision

Refer to the attached worksheet. For each block containing an entry (i.e., performance measurement is possible), your task is to rate the measurement method on the precision of the data it will provide. Measurement precision ratings are a composite of two factors: reliability and validity. The next series of paragraphs describe how you are to provide reliability and validity ratings for the measurement methods associated with DORAC alternatives.

Reliability

Reliability is defined as the extent to which a measurement method provides accurate and stable performance scores. Method reliability ratings are to be assigned using the following procedure:

1. For each performance, order the DORAC alternatives from "best" to "worst" according to the judged reliability of their associated measurement method. Ties are permitted. If one or more of the alternatives are judged equivalent in terms of the reliability of their measurement methods (e.g., they employ the same method), assign them the same rank. Enter ranks in column A.
2. Numerically position the best and worst alternatives on a 0-to-100 scale. Use the following anchor rating points as a guide and enter rating in column B.
   - 0 - The method will provide scores that are completely unreliable.
   - 25 - The method will provide scores having low reliability.
   - 50 - The method will provide scores having a moderate level of reliability.
   - 75 - The method will provide scores having high reliability.
   - 100 - The method will provide scores that are completely reliable.
3. Position the remaining alternatives between the best and worst cases on the 0-to-100 reliability scale and enter ratings in column B. Again, refer to the anchor rating points presented above as a guide.

Validity

Validity is defined as the extent to which an evaluatee's score, as defined through the operational performance measure (OPM), is a true representation of the performance it is intended to measure. Operational performance measure validity ratings are assigned using the following series of steps:

1. For each performance, order the DORAC alternatives from "best" to "worst" according to the judged validity of their associated OPM. Ties are permitted. If one or more (or all) of the alternatives are judged equivalent in terms of the validity of their OPM (i.e., they employ the <u>same</u> OPM), assign them the same rank. Enter rank in column C.
2. Numerically position the best and worst alternatives on a 0-to-100 scale. Use the following anchor rating points as a guide and enter ratings in column D:
   - 0 - The OPM represents a completely invalid operational definition of the performance.
   - 25 - The OPM has <u>low</u> validity with respect to the performance statement. A <u>large portion</u> of the essence of the performance is not reflected in the OPM.
   - 50 - The OPM has <u>moderate</u> validity. Most of the essence of the performance is reflected in the OPM.
   - 75 - The OPM has <u>high</u> validity. Nearly all of the essence of the performance is reflected in the OPM.
   - 100 - The OPM has perfect validity. The complete essence of the performance statement is reflected in the OPM. Essentially, the OPM <u>is</u> the performance.
3. Position the remaining alternatives between the best and worst cases on the 0-to-100 validity scale. Again, refer to the anchor rating points presented above as a guide and enter ratings in column D.

Date

Group #

Job

STEP 3B

Measurement Precision Worksheet

Performance Measure	Alternative 1 — Device: Record Fire; OPM Measurement Method	A	B	C	D	Alternative 2 — Device: Weaponeer; OPM Measurement Method	A	B	C	D
I. Prepare to Engage Targets										
1. Condition Weapon to Operate	Correct impact point; group									
A. Calibrate (zero) Weapon	Evaluator records group					0				
1. Adjust for minute of angle										
2. Adjust for gravity effects	Evaluator records group					0				
3. Obtain proper sight-weapon alinement	Evaluator records group					0				
B. Load Weapon	Go-NoGo									
4. Inspect Magazine for top round alinement	Evaluator Checklist					0				
5. Apply proper intensity flow to seat magazine	Evaluator Checklist					Go-NoGo Evaluator checklist				
6. Insure first round is seated in chamber	Evaluator Checklist					0				

Step 3,4A

Date ____________

Group # ____________

Job ____________

Measurement Precision Worksheet

Performance Measure	Alternative 1: Device Record Fire — OPM Measurement Method	A	B	C	D	Alternative 2: Device Weaponeer — OPM Measurement Method	A	B	C	D
2. Maintain Weapon Operation	Go-NoGo					Go-NoGo				
C. Reload weapon	Evaluator Observation					Evaluator Observation				
7. Remove magazine										
4. Same as above	Evaluator Observation					0				
5. Same as above	Evaluator Observation					Go-NoGo Evaluator Observ.				
6. Same as above	Evaluator Observation					0				
D. Reduce stoppage	Go-NoGo					Go-NoGo				
8. Perform immediate action procedure	Evaluator Checklist					Evaluator Checklist				
9. Inspect weapon and diagnose stoppage	Evaluator Checklist					0				
10. Take appropriate action for reducing stoppage	Evaluator Checklist					0				
3. Operate Weapon	Go-NoGo					Go-NoGo				
E. Soldier-weapon Interface	Evaluator Checklist					Evaluator Checklist Machine assisted				
11. Trigger control (squeeze)										
12. Grip on weapon	Evaluator Checklist					Evaluator Checklist				
13. Breath control	Evaluator Checklist					Evaluator Checklist				
14. Relaxation	Evaluator Checklist					Evaluator Checklist				

(Continued)

Step 3,4A

Date ______

Group # ______

Job ______

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device Record Fire OPM Measurement Method	A	B	C	D	Alternative 2 Device Weaponeer OPM Measurement Method	A	B	C	D
15. Compensate for recoil	Evaluator Checklist					Evaluator Checklist				
16. Stock weld	Evaluator Checklist					Evaluator Checklist				
17. Firing elbow	Evaluator Checklist					Evaluator Checklist				
F. Marksmanship Factors 18. Compensate for range	Subjective Go-NoGo Observer records score for each target					Subjective Go-NoGo Evaluator verifies machine assisted				
19. Compensate for wind	Observer records score for each target					0				
20. Compensate for target movement	0					0				
21. Conduct transfer of fire	Subjective Go-NoGo Observer records score/behavior					Subjective Go-NoGo Observer records score/behavior				
22. Obtain proper sight picture	Go-NoGo M-16 sighting device					Go-NoGo M-16 sighting device, video playback				
23. Determine proper aim point	Subjective Go-NoGo Evaluator observes hit pattern					Go-NoGo Machine video playback				
24. Determine proper weapon aspect-round impact relationship	Subjective Go-NoGo Evaluator observes behavior					0				

Step 3,4A

Date ____________

Group # ____________

Job ____________

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device OPM: Record Fire Measurement Method	A	B	C	D	Alternative 2 Device OPM: Weaponeer Measurement Method	A	B	C	D
II. Engage Targets (In): 4. Assault Mode (using) g. Aimed fire (from) 25. Prone position	Minimum 17 hits/40 rounds (23 hits at Ft. Benning) Observer records hits (composite of a,c,&d)					Predetermined percentage of hits (variable) Machine recorded score printout, video playback				
26. Standing position	Not scored					"				
27. Foxhole supported position	See (a)					"				
28. Foxhole unsupported position	See (a)					"				
h. Pointed fire: 29. From standing position	Not currently assessed Observer records hits (composite of e & f)					0				
30. While moving	See (e)					0				
5. Defensive positions (using): g. Aimed fire (from) 25. Prone position	Same as 25 above)					(Same as 25 above)				
26. Standing position	0					"				
27. Foxhole supported position	See (a)					"				

Step 3,4A

Date ____________

Group # ____________

Job ____________

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device Record Fire OPM Measurement Method	A	B	C	D	Alternative 2 Device Weaponeer OPM Measurement Method	A	B	C	D
28. Foxhole unsupported position	See (a)					(Same as 25 above)				
h. Pointed fire: 29. From standing position	(Same as 26 above)					0				
30. While moving	(See (e)					0				
6. Patrol Operation (using): g. Aimed fire (from): 25. Prone position	Minimum 17 hits/40 rounds (23 hits at Ft. Benning) Observer records hits (composite of a,c,&d)					Predetermined percentage of hits (variable) Machine recorded score Printout, video playback				
26. Standing position	Not scored					"				
27. Foxhole supported position	See (a)					"				
28. Foxhole unsupported position	See (a)					"				
h. Pointed fire: 28. From standing position	Not currently assessed Observer records hits (composite of e & f)					0				
29. While moving	See (e)					0				

Step 3,4A

Date ______________

Group # ______________

Job ______________

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device Record Fire OPM Measurement Method	A	B	C	D	Alternative 2 Device Weaponeer OPM Measurement Method	A	B	C	D
7. Detect and Identify Targets	0					0				
III. Team Fire.										
8. Conduct fire and maneuver i. Control fire team movement	0					0				
j. Move as a member of a fire team	0					0				
k. Move under direct fire	0					0				
9. Practice ammunition rationing l. Control rate of fire	0					0				
m. Control distribution of fire	0					0				

Step 3,4A

STEP 4A

Effectiveness Ratings

System effectiveness is defined as the degree to which a DORAC device provides timely, quality information on the performances under consideration. Specifying system effectiveness is carried out in two steps: First, information quality ratings are obtained for each device on each performance. Second, each DORAC device is evaluated with respect to the utility of the frequency with which performance data are provided. The next series of paragraphs describe the effectiveness rating procedure in additional detail. Please use the form on the next page to record your results.

Information Quality

Information quality is defined as the extent to which a device provides precise information relevant to a particular performance. Also considered as part of information quality is the amount of information provided by a device; that is, the number of relevant performance conditions that are addressed by the device. Information quality ratings are obtained using the procedure outlined as follows:

1. For each performance, order the DORAC devices from "best" to "worst" according to the degree to which the devices are capable of providing quality information relevant to the performance under consideration. Factors that should be considered in making quality judgments include:
   a. Amount of information. The number of relevant performance conditions that are addressed.
   b. Precision. The judged precision of the data. This is obtained from the measurement precision ratings assigned previously.

   Ties are permitted. If one or more of the devices are judged equivalent in terms of the quality of the information they provide, assign them the same rank. Record ranks in column A.

2. Numerically position the best and worst devices on a 0-to-100 scale in column B. Use the following anchor points as a guide:

   0 - The device provides no data relevant to the performance under consideration.

   25 - Marginal. The device provides partial data on the performance and the recording/scoring method is poor resulting in low validity or low reliability.

   50 - Adequate. The device provides required data, but some measurement precision problems are apparent. For example, the most appropriate recording/scoring method is not used or the data is likely to have only moderate reliability.

   75 - Good. The device provides required data in an acceptable manner. Recording methods are acceptable; reliability is likely to be quite high.

   100 - Excellent. The device is the best possible, given the current technical state of the art. Recording methods are automated and precise; reliability is likely to be very high.

3. Position the remaining devices between the best and worth cases on the 0-to-100 scale and enter ratings in column B. Again, refer to the anchor rating points presented above as a guide.

Frequency Utility

The second step in obtaining effectiveness ratings is to determine the utility of the evaluation frequency associated with each alternative. Frequency utility ratings are obtained by applying the following sequence of actions:

1. Consider the frequency of the information provided by each DORAC device (e.g., quarterly, twice a year, yearly, etc.). Now, specifically considering the highest and lowest evaluation frequencies, rate the utility of receiving performance status information with these frequencies. Use a 0-to-100 scale in assigning your utility ratings.

2. Next, consider the remaining alternatives. Position the utility of their evaluation frequencies between the extreme values (i.e., ratings for the highest and lowest frequencies) on the 0-to-100 scale.

Date ____________

Group # ____________

Job ____________

STEP 4A

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device Record Fire OPM Measurement Method	A	B	C	D	Alternative 2 Device Weaponeer OPM Measurement Method	A	B	C	D
I. Prepare to Engage Targets 1. Condition Weapon to Operate	Correct impact point; group									
A. Calibrate (zero) Weapon 1. Adjust for minute of angle	Evaluator records group					0				
2. Adjust for gravity effects	Evaluator records group					0				
3. Obtain proper sight-weapon alinement	Evaluator records group					0				
B. Load Weapon 4. Inspect Magazine for top round alinement	Go-NoGo Evaluator Checklist					0				
5. Apply proper intensity flow to seat magazine	Evaluator Checklist					Go-NoGo Evaluator checklist				
6. Insure first round is seated in chamber	Evaluator Checklist					0				

Step 3,4A

Date ______________

Group # ______________

Job ______________

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device Record Fire OPM Measurement Method	A	B	C	D	Alternative 2 Device Weaponeer OPM Measurement Method	A	B	C	D
2. Maintain Weapon Operation	Go-NoGo					Go-NoGo				
C. Reload weapon	Evaluator Observation					Evaluator Observation				
7. Remove magazine										
4. Same as above	Evaluator Observation					0				
5. Same as above	Evaluator Observation					Go-NoGo Evaluator Observ.				
6. Same as above	Evaluator Observation					0				
D. Reduce stoppage	Go-NoGo					Go-NoGo				
8. Perform immediate action procedure	Evaluator Checklist					Evaluator Checklist				
9. Inspect weapon and diagnose stoppage	Evaluator Checklist					0				
10. Take appropriate action for reducing stoppage	Evaluator Checklist					0				
3. Operate Weapon	Go-NoGo					Go-NoGo				
E. Soldier-weapon interface	Evaluator Checklist					Evaluator Checklist Machine assisted				
11. Trigger control (squeeze)										
12. Grip on weapon	Evaluator Checklist					Evaluator Checklist				
13. Breath control	Evaluator Checklist					Evaluator Checklist				
14. Relaxation	Evaluator Checklist					Evaluator Checklist				

(Continued)

Step 3,4A

Date ____________________

Group # ____________________

Job ____________________

Measurement Precision Worksheet

	Alternative 1					Alternative 2				
	Device Record Fire					Device Weaponeer				
Performance Measure	OPM Measurement Method	A	B	C	D	OPM Measurement Method	A	B	C	D
15. Compensate for recoil	Evaluator Checklist					Evaluator Checklist				
16. Stock weld	Evaluator Checklist					Evaluator Checklist				
17. Firing elbow	Evaluator Checklist					Evaluator Checklist				
F. Marksmanship Factors 18. Compensate for range	Subjective Go-NoGo Observer records score for each target					Subjective Go-NoGo Evaluator verifies machine assisted				
19. Compensate for wind	Observer records score for each target					0				
20. Compensate for target movement	0					0				
21. Conduct transfer of fire	Subjective Go-NoGo Observer records score/behavior					Subjective Go-NoGo Observer records score/behavior				
22. Obtain proper sight picture	Go-NoGo M-16 sighting device					Go-NoGo M-16 sighting device, video playback				
23. Determine proper aim point	Subjective Go-NoGo Evaluator observes hit pattern					Go-NoGo Machine video playback				
24. Determine proper weapon aspect-round impact relationship	Subjective Go-NoGo Evaluator observes behavior					0				

Step 3,4A

Date ______________

Group # ______________

Job ______________

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device Record Fire OPM Measurement Method	A	B	C	D	Alternative 2 Device Weaponeer OPM Measurement Method	A	B	C	D
II. Engage Targets (In): 4. Assault Mode (using) g. Aimed fire (from) 25. Prone position	Minimum 17 hits/40 rounds (23 hits at Ft. Benning) Observer records hits (composite of a,c,&d)					Predetermined percentage of hits (variable) Machine recorded score printout, video playback				
26. Standing position	Not scored					"				
27. Foxhole supported position	See (a)					"				
28. Foxhole unsupported position	See (a)					"				
h. Pointed fire: 29. From standing position	Not currently assessed Observer records hits (composite of e & f)					0				
30. While moving	See (e)					0				
5. Defensive positions (using): g. Aimed fire (from) 25. Prone position	Same as 25 above)					(Same as 25 above)				
26. Standing position	0					"				
27. Foxhole supported position	See (a)					"				

Step 3,4A

Date ______

Group # ______

Job ______

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device Record Fire OPM Measurement Method	A	B	C	D	Alternative 2 Device Weaponeer OPM Measurement Method	A	B	C	D
28. Foxhole unsupported position	See (a)					(Same as 25 above)				
h. Pointed fire: 29. From standing position	(Same as 26 above)					0				
30. While moving	(See (e)					0				
6. Patrol Operation (using): g. Aimed fire (from): 25. Prone position	Minimum 17 hits/40 rounds (23 hits at Ft. Benning) Observer records hits (composite of a,c,&d)					Predetermined percentage of hits (variable) Machine recorded score Printout, video playback				
26. Standing position	Not scored					"				
27. Foxhole supported position	See (a)					"				
28. Foxhole unsupported position	See (a)					"				
h. Pointed fire: 28. From standing position	Not currently assessed Observer records hits (composite of e & f)					0				
29. While moving	See (e)					0				

Step 3,4A

Date ____________

Group # ____________

Job ____________

Measurement Precision Worksheet

Performance Measure	Alternative 1 Device Record Fire OPM Measurement Method	A	B	C	D	Alternative 2 Device Weaponeer OPM Measurement Method	A	B	C	D
7. Detect and Identify Targets	0					0				
III. Team Fire.										
8. Conduct fire and maneuver i. Control fire team movement	0					0				
j. Move as a member of a fire team	0					0				
k. Move under direct fire	0					0				
9. Practice ammunition rationing										
l. Control rate of fire	0					0				
m. Control distribution of fire	0					0				

Step 3,4A

STEP 4B

Effectiveness Ratings

System effectiveness is defined as the degree to which a DORAC device provides precise, timely information on all aspects of the performances under consideration. Specifying system effectiveness is carried out in two steps. First, ratings are obtained regarding the importance of each of the condition variables. When these ratings are combined with the condition variable coverage capabilities of each of the DORAC alternatives, a system capability rating is obtained. You will not, however, have to perform this latter action.

The second step involves obtaining evaluation frequency utility ratings. Frequency utility ratings will be assigned by considering the decay rate of each performance. These two factors--coverage of condition variables and frequency utility--are combined with the Measurement Precision ratings you provided earlier to obtain an effectiveness rating for each DORAC device on each performance. Again, you will not have to combine the ratings. The next series of paragraphs describe the effectiveness rating procedure in additional detail. For your convenience, appropriate rating forms are provided.

Importance of Condition Variables

Consider each of the condition variables listed in column A on the attached rating sheet. Assign importance ratings to each of the condition variables using the following steps:

1. Rank the condition variables in order of their importance for consideration in performance measurement in column B. If sub-condition variables are nested under a given condition variable, rank them in order of their importance relative to that specific sub-set on a separate rating sheet.
2. Assign the least important condition variable a rating of 10 in column C.
3. Consider the next-least-important condition variable. How much more important is it than the least important? Assign it a number that relfects that ratio. For example, if the second-least-important condition variable is judged to be four times as important as the first, it is assigned a score of 40. Continue up through the list of condition variables. Check each set of ratios as each new judgment is made and enter in column C.

4. Review your ratings to insure that they reflect the actual importance of each of the condition variables. Are the ratios of distances between condition variables correct? Make any necessary adjustments to your ratings in column D.
5. If sub-condition variables are nested under a given condition variable, repeat Steps 2 through 4 on the items within each sub-set. Remember, assign ratings to each sub-set individually.

Frequency Utility

Now, consider the range of evaluation frequencies associated with each of the DORAC alternatives (e.g., one, two, three times per year, etc.). Using the worksheet provided, identify the minimum and maximum evaluation frequencies; fill in all values between these two extremes. Now, specifically consider as a group the performances rated as having a low ("L") decay rate. Assign frequency utility ratings to each of the actual and potential evaluation frequencies for performances rated "L" by applying the following sequence of actions:

1. Focus on the highest and lowest evaluation frequencies. Rate the utility or usefulness of receiving proficiency status information with the frequencies indicated. Use a 0-to-100 scale in assigning your ratings.
2. Now, consider the intermediate evaluation frequencies. Position the remaining frequencies between the extreme values (i.e., the ratings for the highest and lowest evaluation frequencies) on the 0-to-100 scale.
3. Create a frequency utility curve by connecting the scaled points with a line. Connect the point associated with the lowest frequency with the zero-point on the Frequency-Utility axes.

Repeat Steps 1 through 3 for the "M" and "H" performance decay rate categories. Place all three frequency utility curves on the same graph. Review the relationships among the three curves. If you are not satisfied with what you observe, go back and adjust the utility ratings until you are satisfied.

Date ______________

Group # ______________

Job ______________

STEP 4B

CONDITION AND TARGET VARIABLES

Rating Development Sheet

A	B	C	D	E										F
				Revised Rating (Part 2)										
Elements	Rank	Initial Rating	Final Rating	1	2	3	4	5	6	7	8	9	10	Importance Weight
1. Multiple Targets														
2. Friendly and Hostile Targets														
3. Variable Range														
4. Target Movement: a. Direction (see b. Distance attached c. Rate sheet)														
5. Target Exposure: a. Amount (see b. Duration attached c. Frequency sheet)														
6. Target Camouflage														
7. Target Termination When Hit														
8. Target-Background Contrast Ratio														
9. Terrain Features														
10. Target Illumination														

Step 1,2,4B

Date ______________

Group # ______________

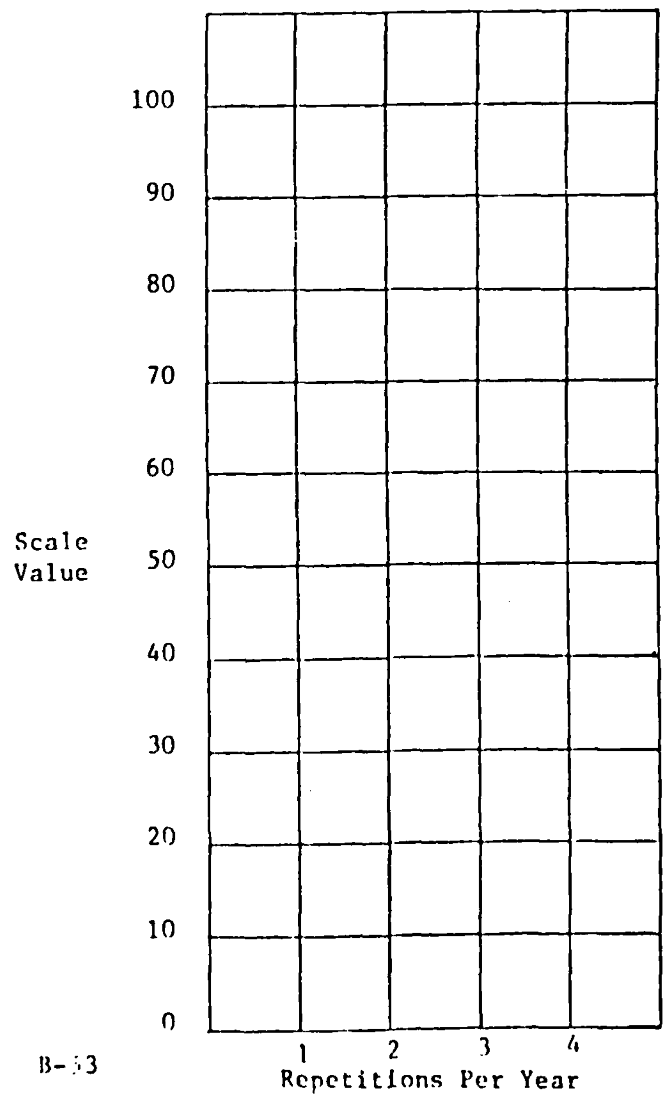
Job ______________

STEP 4B

Rating Development Sheet

A	B	C	D	E										F
				Revised Rating (Part 2)										
Elements	Rank	Initial Rating	Final Rating	1	2	3	4	5	6	7	8	9	10	Importance Weight
TARGET MOVEMENT														
a. Direction														
b. Distance														
c. Rate														
TARGET EXPOSURE														
a. Amount														
b. Duration														
c. Frequency														

Step 1,2,4B


Scale
Value
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
Repetitions Per Year

APPENDIX C

Procedure for Weighting Information Worth Dimensions

STEP 1

Importance Weights for Worth Dimensions

Part 1:

Importance weights for Worth Dimensions (WDs) are assinged using the series of steps presented below. To assist in the rating process, a rating development sheet is provided on the next page.

1. Rank the WDs in order of importance in column B.
2. Rate the WDs on importance: (Column C)
   a. Assign the least important WD a rating of 10 in column C.
   b. Consider the next-least-important WD. How much more important is it than the least important? Assign it a number that reflects that ratio. For example, if the second-least-important WD is judged to be four times as important as the first, it is assigned a score of 40. Continue up through the list of WDs. Check each set of ratios as each new judgment is made.
   c. Review your ratings to insure that they reflect the actual importance of each of the WDs. Are the ratios of distances between WDs correct? Make any necessary adjustments in your ratings and list the results in column F.

Part 2:

A. If only two WDs are noted, sum the resulting scores. Divide each score by the resulting sum. Round to two places. Record results in column F, which completes this step.

B. If more than two (2) WDs are being rated, carry out the following additional series of steps to improve the reliability of the resulting importance weights, using column E, 1 to 10, for each repetition.

1. Compare the first (most important) WD with the remaining ones put together. Is it more important, equally important, or less important than all the others put together?
2. If the first WD is more important than all of the others put together, see if it's importance rating is greater than the sum of the importance ratings for all of the other WDs. If not, change the importance rating of the first WD so that it is greater than the sum of the others.

3. If the first WD is of equal importance to all the others put together, see if its importance rating is equal to the sum of the importance ratings of all the other WDs. If it is not, change the importance rating of the first WD so that it is equal to the sum of the others.
4. If the first WD is less important than all the others put together, see if its importance rating is less than the sum of the importance ratings of all of the other WDs. If it is not, change the importance rating of the first WD so that it is less than the sum of the others.
5. If the first WD was considered more important or equally important than all the others put together, apply the above procedure to the second-most-important WD on the list. Is it more important, equally important, or less important than all the other farther down the list put together? Then, proceed as in (2), (3), and (4) above, applying the revision procedure to the second WD instead of the first.
6. If the first WD was considered less important than all the others put together, compare the first WD with all the remaining ones put together, except the lowest rated one. Is the first WD more important, equally important, or less important than all of the others farther down the list except the lowest one put together? Then proceed as in (2), (3), and (4) above. If (2) or (3) are applicable, proceed to (5) after applying (2) or (3). If (4) is applicable, proceed as in this paragraph (6) again, comparing the first WD with all the remaining ones put together except the lowest two. As long as (4) is applicable, the procedures of this paragraph (6) are repeated until the first WD is compared with the second and third WDs put together. Then, even if (4) is still applicable, proceed to (5).
7. Continue the above procedure until the third-from-the-lowest WD has been compared with the two lowest WDs on the list.
8. Sum the resulting scores. Divide each score by the resulting sum. Round to two places. Record results in column F, which completes this step.

APPENDIX D

Performance Context Variable List

Condition Variable Network

Task Controlling Stimuli (General)	visual stimuli	- unprocessed real world	- equipment related - external environment peculiar
		- processed video	
	auditory	- voice - encoded - equipment produced - ambient	
	tactile	- control movement - shape - temperature	
	olfactory	- equipment produced smells - object peculiar smells	
	motion	- internal motion cues (tactile?) - external motion cues (visual?) - motion of equipment (visual?)	

Target Variables (Specific Task Controlling Stimuli)	multiple targets	
	target identification	
	variable target range	
	target movement	distance direction rate
	target exposure	amount duration frequency
	target signature	camouflage target/background contrast ratio target illumination
	target termination	
	terrain features	
	target counterfire	
	target scale	exact scale reduced scale enlarged scale variable

Condition Variables	Performance Context	individual
		team (non-interactive)
		team (interactive)
	Weather	wind
		precipitation
		temperature
		dust
	Visible light	daylight
		moonlight
		white searchlight
		infrared searchlight
		moonless darkness
		obscuration
	Other	motion
		space limitations

APPENDIX E

Guidelines for Obtaining Measurement Precision Ratings

Step 3. Evaluation of Measurement Precision

In this step, you are going to make evaluations of the CEV gunnery training devices proposed as DORACs. You will rate how well the different devices can measure a given performance. These ratings are important because devices often differ in their ability to evaluate performances. Five training devices and evaluation systems are being considered as CEV gunnery DORACs. They are:

1. Range fire--a live fire qualification with the maingun and machineguns.
2. CEV crew gunnery skills test--the test of gunnery skills defined in FM 17-12-6, to be completed prior to range fire.
3. The Saunders Interactive Video Tape System (IVTS)--a video tape driven device that trains the Gunner in main gun engagements (primarily a BOT trainer).
4. The improved IVTS--similar to the IVTS only the device is driven by video disc, machine gun engagements also can be performed, boresighting can be performed to some extent, and some crew interaction is possible.
5. The Perceptronics device--essentially the same as the improved IVTS.

You should be familiar with the different devices and evaluation systems. If you have any questions, the ASA representative will answer them.

When we talk about how well a device can evaluate a performance we are really asking several questions about the measure the device provides of the performance. First, we must ask how much the performance itself is actually assessed? Consider the task of engaging a target with an M-16 rifle. One means of assessing a soldier's ability to engage targets with the M-16 is to provide the soldier with an M-16 and have him engage targets. Another method of assessment would be to give the soldier a multiple choice test on how to engage targets with the M-16. Most people would agree that actual operation of the M-16 provides a better means of assessment than the multiple choice test. Generally, the more a device requires or simulates the actual performance to be assessed, the better its ability to evaluate the performance.

A second question, related to the first, is how much of a performance is actually required by a device? Let's say we want to evaluate a soldier's ability to clear a jammed round from the M-16. One means of evaluating this task would be to arrange for a round to jam in the soldier's weapon and observe the procedures he follows to clear the weapon. Another means of evaluation would be to use the Weaponeer, a trainer for the M-16. The Weaponeer can simulate the occurrence of a jam (weapon does not fire), but its abilities to assess clearning procedures are limited. To clear the simulated jam, the soldier simply charges the weapon once and resumes firing. Thus, the Weaponeer does provide some assessment of a soldier's ability to clear the M-16, but it does not provide as good an evaluation as observing a soldier with an actual jam in a weapon. Generally, the more a device requires all elements of a performance the better the evaluation it provides.

The first two questions we discussed were concerned with what is called the <u>validity</u> of a measure. The third question is concerned with a measure's <u>reliability</u>. By reliability, we mean how accurate and consistent is the device in measuring a performance? Accuracy refers to the precision of the measurement system. A device that measures the percent of task steps performed correctly is more accurate than one that simply gives a satisfactory/unsatisfactory evaluation. Consistency refers to a device's ability to give the same evaluation of repeats of the same performance. For example, if two soldiers fire at targets and both hit nine out of ten, does the device score both with 90% hits? If so, it is consistent in its evaluation. When humans do the evaluating, a good way to think of consistency is in terms of agreement. Ask yourself, "If I had 10 of my people evaluate a soldier's performance, would they all come up with the same score or evaluation of that performance? Generally, machines are considered to be more reliable at evaluating performances. However, when a performance is not too complex or does not occur too rapidly humans can be very reliable evaluators.

Ratings of DORAC Measurement Effectiveness

A data sheet is attached to record your ratings. The data sheet identifies the devices capable of measuring each performance and gives a brief description of the measure used by the device.

Step 1. For a given performance, study the measures used by the devices that can assess the performance. Rank the devices in terms of their ability to measure the performance effectively (if only one device can assess the performance give it a rank of one [1]).

Step 2. Consider the device ranked as providing the most effective measure of the performance. Using a 0 to 100 scale, rate the measurement effectiveness of the device. Use the following as reference points on the scale:

- 0 - does not evaluate the performance at all, is completely inaccurate, or totally inconsistent.
- 25 - provides some evaluation (e.g., perhaps is inaccurate or inconsistent, does not assess the actual performance).
- 50 - provides a moderately good evaluation (e.g., assesses part of the actual performance or lacks the accuracy or consistency desired).
- 75 - provides a good evaluation (e.g., most of the actual performance is required, the scoring system is reasonably accurate and consistent).
- 100 - is a perfect evaluation system, requires actual, realistic performance of the task, has the accuracy and consistency desired.

If there is only one device capable of measuring this performance, proceed to the next performance and begin with Step 1.

Step 3. Using the same scale, rate the measurement effectiveness of the device ranked as the least effective measurement method.

Step 4. Rate the measurement effectiveness of the remaining devices, if any.

Step 5. Proceed to the next performance and repeat Steps 1-4.

APPENDIX F

Guidelines for Obtaining Frequency Utility Ratings

Step 5

Frequency Utility Ratings

1. When an evaluation is set up, one of the key questions to be answered is, how often do you evaluate people? In order to make this decision for a DORAC, information about frequency utility (the usefulness of different lengths of time between evaluations) must be gathered. In this step you will provide information about the value of receiving evaluation information once, twice, three times, and four times a year. This must be done for both possible uses for the information, Unit Readiness Evaluation, and Unit Training Management. The value of evaluation frequencies will be rated for each task cluster under each Worth Dimension. The results of this step will help to select the best evaluation frequency of a DORAC for your system.

2. When rating evaluation frequencies, some key points must be kept in mind. First, some types of tasks need more frequent practice to stay proficient than others. Tasks that have a high decay rate (that is, people rapidly fall below standards without frequent practice) may need to be evaluated more frequently than tasks with a moderate or low decay rate. Moderate decay rate tasks need to be practiced every three or four months to meet standards. Low decay rate tasks, once learned, need little or no practice. The second thing to consider is that the frequency of evaluation really expresses the possible age of the evaluation information. Once a year evaluation means that information is a year old before the next evaluation. Twice a year gives you information that can be six months old. Three times a year gives four months old information, and four times a year would be three months old before the next evaluation. A third factor to consider is that people change with time. When your type of unit has a high turnover rate, few of the people evaluated are still there after a year. The value of evaluation information is reduced when the people evaluated are not longer in the unit. A fourth factor is that evaluations that are conducted too frequently may reduce the information value if personnel morale is affected by the effort required to prepare and conduct evaluations, or if

they become bored or overtrained to the point that they don't give their best effort for evaluations. When rating frequency utility, keep in mind that different types of tasks are forgotten at different rates, that evaluation frequencies relate to how old information may be when you use it, that personnel turnover affects information value as it ages, and that it is possible to evaluate too frequently.

Your job for this step is to consider, for each worth dimension, keeping in mind decay rates, turnover rates, and the possibility of "overtraining", for the worst case for each frequency (i.e., 12 month old information for once a year) what percentage of confidence or trust you could place on the evaluation information. In this step you will be using a 0 to 100 scale to indicate your degree of confidence.

STEP 5

Ratings of Frequency Utility

A 0 to 100 scale will be used to rate the utility of the different evaluation frequencies. The following anchor points on the scale are provided to aid you in making ratings.

0 - information received at the frequency under consideration has no value at all (e.g., evaluation occurs infrequently and the tasks evaluated are forgotten quickly or there is a very high turnover rate in the unit. You would not have an accurate picture of unit status shortly after the evaluation occurred)

25 - information received at the frequency under consideration has some value but not much (e.g., unit turnover might be high but the tasks evaluated are not forgotten too quickly. Evaluation occurs often enough to give you some idea of the unit's status)

50 - information received at the frequency under consideration has moderate value (e.g., much needed information is provided, however, evaluations are conducted so frequently that over-training has occurred and unit morale is affected. The value of the information gained is offset by the decline in morale)

75 - information received at the frequency under consideration has high value but not maximum value (e.g., given the decay rates of tasks evaluated and turnover in the unit you feel that a pretty good picture of unit status is obtained under the evaluation frequency. Perhaps evaluations are conducted a bit more than you would like)

100 - information received at the frequency under consideration has the maximum or greatest possible value. A very accurate picture of unit status is maintained and evaluations are not conducted too frequently.

Of course, your actual ratings may fall anywhere between 0 and 100. These points are provided only to establish a scale to aid you in making rating decisions.

The first part of this step is to rate the frequency utility for evaluation information that will be used to determine unit readiness. Using the scale described above, consider the assessment of unit readiness and rate the worth for assessing readiness of receiving information on all CEV gunnery tasks. Your ratings must represent your best estimate of the value to you of information (what percentage of confidence in its accuracy) for each frequency time period.

1. Once a year (every twelve months) rating: ____________
2. Twice a year (every six months) rating: ____________
3. Three times a year (every four months) rating: ____________
4. Four times a year (every three months) rating: ____________

Continue to consider the assessment of unit readiness and rate the value for assessing readiness of receiving information on tasks in each of the major clusters of CEV gunnery tasks at the different evaluation frequencies.

Frequency utility of performance information on Prepare to Fire tasks:

1. Value of information provided once a year (every twelve months) rating: ____________
2. Value of information provided twice a year (every six months) rating: ____________
3. Value of information provided three times a year (every four months) rating: ____________
4. Value of information provided four times a year (every three months) rating: ____________

Frequency utility of performance information on Ammunition Handling tasks:

1. Value of information provided once a year rating: ____________
2. Value of information provided twice a year rating: ____________
3. Value of information provided three times a year rating: ____________
4. Value of information provided four times a year rating: ____________

Frequency utility of performance information on Occupy Firing Position tasks:

1. Value of information provided once a year. rating: ________
2. Value of information provided twice a year. rating: ________
3. Value of information provided three times a year. rating: ________
4. Value of information provided four times a year. rating: ________

Frequency utility of performance information on Target Engagement tasks:

1. Value of information provided once a year. rating: ________
2. Value of information provided twice a year. rating: ________
3. Value of information provided three times a year. rating: ________
4. Value of information provided four times a year. rating: ________

The second part of this step concerns Worth Dimension two, Unit Training Management. Considering the management of unit training, rate the worth for managing unit training of receiving information on all CEV gunnery tasks.

1. Once a year. rating: ________
2. Twice a year. rating: ________
3. Three times a year. rating: ________
4. Four times a year. rating: ________

Continue to consider the assessment of training management and rate the value for managing unit training of receiving information on tasks in each of the major clusters of CEV gunnery tasks at the different evaluation frequencies.

Frequency utility of performance information on Prepare to Fire tasks:

1. Value of information provided once a year. rating: ________
2. Value of information provided twice a year. rating: ________
3. Value of information provided three times a year. rating: ________
4. Value of information provided four times a year. rating: ________

Frequency utility of performance information on Ammunition Handling tasks:

1. Value of information provided once a year. rating: ________
2. Value of information provided twice a year. rating: ________
3. Value of information provided three times a year. rating: ________
4. Value of information provided four times a year. rating: ________

Frequency utility of performance information on Occupy Firing Position tasks:

1. Value of information provided once a year. rating: ________
2. Value of information provided twice a year. rating: ________
3. Value of information provided three times a year. rating: ________
4. Value of information provided four times a year. rating: ________

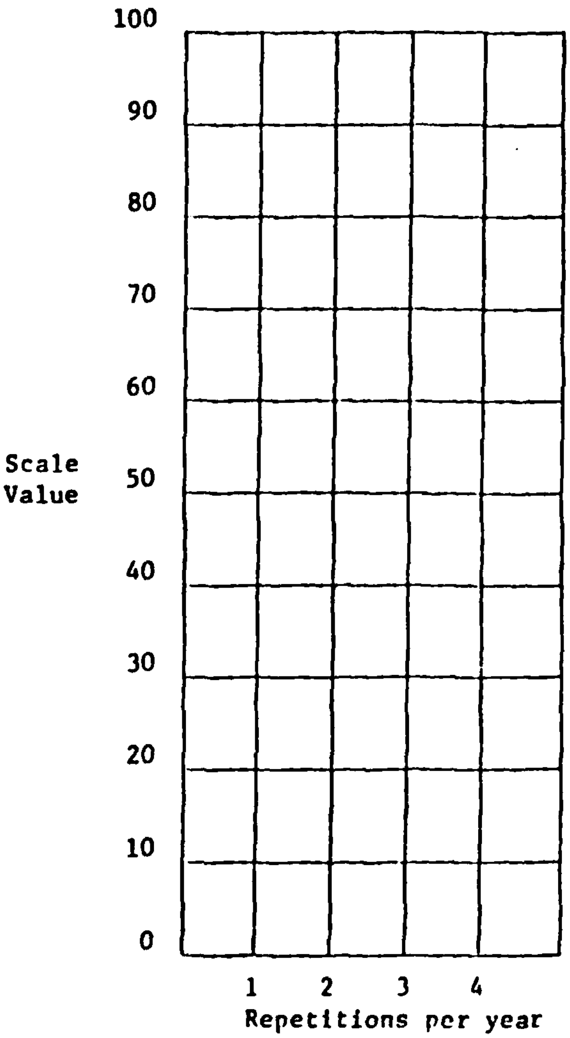
Frequency utility of performance information on Target Engagement tasks:

1. Value of information provided once a year. rating: ________
2. Value of information provided twice a year. rating: ________
3. Value of information provided three times a year. rating: ________
4. Value of information provided four times a year. rating: ________

Now plot out ratings, in pencil, for each worth dimension on the graphs provided on the next page. Use one symbol to indicate rating for each frequency for one set of ratings, and then connect symbols with a pencil line. Plot all five sets of ratings for each worth dimension. Then compare the plot lines. If you are satisfied with the results, this step is completed. If you are not satisfied, make necessary corrections on rating pages and graphs.

Symbols: All gunnery tasks = ●
Prepare to Fire tasks = ○
Ammunition Handling tasks = □
Occupy Firing Position tasks = ◇
Target Engagement tasks = ☆

WD#1
Unit Readiness Assessment



WD#2
Unit Training Management